

MAINTENANCE OF GERM LINE AND SOMATIC DNA METHYLATION DURING
MOUSE DEVELOPMENT

by

Bonnie Lynn Reinhart

BS, The Pennsylvania State University, 1997

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2003

UNIVERSITY OF PITTSBURGH
FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Bonnie Lynn Reinhart

It was defended on

August 18, 2003

and approved by

Dr. Gregg Homanics

Dr. Jeffrey Hildebrand

Dr. Karen Arndt

Dr. Deborah Chapman
Committee Co-Chairperson

Dr. John Richard Chaillet
Committee Co-Chairperson

MAINTENANCE OF GERM LINE AND SOMATIC DNA METHYLATION DURING MOUSE DEVELOPMENT

Bonnie Lynn Reinhart, PhD

University of Pittsburgh, 2003

DNA methylation in mammals is involved in several essential processes including X chromosome inactivation, genomic imprinting, and host defense against mobile genetic elements. How methylation is targeted to specific sequences in the germ line and how specific methylation patterns are maintained during development is not fully understood. Genomic methylation is established in the gamete, decreases dramatically during preimplantation development, and is re-established after implantation of the blastocyst. However, the methylation present at imprinted loci is specifically maintained during preimplantation development. Imprinted genes are located in clusters, and within each gene cluster parent-of-origin specific expression is governed by an imprinting center (IC). The ICs of the maternally imprinted murine *Snrpn*, *Kcnq1*, and *Igf2r* gene clusters coincide with their differentially methylated domains (DMDs). We have shown that specific DMD sequences are required to establish differential methylation at an imprinted locus. Hybrid transgenes were generated using a non-imprinted derivative of the imprinted *RSV Igmyc* mouse transgene, *Ig/myc*, and sequences from endogenous imprinted gene DMDs. Addition of specific DMD sequences to the *Ig/myc* transgene restored its imprinting. Only the tandem repeats found within the *Snrpn*, *Kcnq1*, and *Igf2r* DMDs were capable of establishing maternal-specific transgene methylation. These experiments have also shown that the methylation on imprinted gene DMD sequences is specifically maintained during the early stages of preimplantation development. These results

clearly demonstrate the importance of tandem repeats in the process of genomic imprinting. DNA methylation is also critical for silencing intracisternal A particle (IAP) transposition within the genome. It is thought that maintenance of IAP element methylation during preimplantation is critical to repress IAP element transcription and transposition. The methylation of IAP element long terminal repeat (LTR) sequences was analyzed at the blastocyst stage of preimplantation development using the bisulfite genomic sequencing technique. These experiments have shown that methylation is maintained on the majority of IAP elements at the blastocyst stage of preimplantation development. However, the methylation on specific IAP elements is completely lost at this time, including the methylation of single IAP element LTRs.

ACKNOWLEDGMENTS

I am extremely grateful to Dr. Richard Chaillet for giving me the opportunity to work in his lab. I have enjoyed the experience, and he has been a wonderful advisor. He has always been a source of ideas for every project I have pursued and I have learned a significant amount during my time in his lab. I would also like to thank my committee members; Deborah Chapman, Karen Arndt, Gregg Homanics, and Jeffrey Hildebrand for everything that they have done to help me along the way.

Thank you to all of the members of the Chaillet lab and the Arndt Lab who have made the lab a wonderful place to work; particularly Sarayu Ratnam, Kathryn Kumer, and Carina Howell who have become great friends.

A special thank you to my family and friends for their continued love and support. I would particularly like to acknowledge my parents, Deborah and William Reinhart, and my brother Bill, they have been behind me in everything that I have ever done, and I could not have achieved any of this without them. Also, a special thanks to Shawn Knight for his unfailing confidence in my ability to do whatever I set my mind to.

TABLE OF CONTENTS

Chapter 1: Introduction	1
1.1. Functions of DNA methylation in mammals.....	1
1.2. Establishment and maintenance of DNA methylation patterns.....	2
1.3. The mammalian DNA methyltransferases	8
1.3.1. DNA methyltransferase 1 activity.....	8
1.3.2. Activity of the Dnmt3 family of methyltransferases.....	13
1.3.3. Possible involvement of Dnmt2 in DNA methylation	14
1.3.4. Interactions between the DNA methyltransferase proteins	15
1.3.5. Other proteins required for genomic DNA methylation.....	15
1.4. DNA methylation and the regulation of gene transcription	16
1.4.1. Methyl-binding domain proteins and transcriptional control	17
1.4.2. Methylation-sensitive DNA binding factors.....	18
1.4.3. DNA methyltransferase interacting factors	18
1.5. Genomic imprinting and DNA methylation.....	19
1.5.1. Genomic imprinting	19
1.5.2. Importance of genomic imprinting.....	20
1.5.3. Genomic imprinting in the mouse.....	21
1.5.4. DNA methylation and genomic imprinting	25
1.5.5. Specific methyltransferases establish and maintain imprinted gene methylation.....	30
1.5.6. Characteristics of imprinted gene differentially methylated domains.....	35
1.6. Silencing of intracisternal A particles in the mouse genome	41
1.6.1. Mobile genetic elements	41
1.6.2. Methylation and control of IAP element transcription.....	47
1.6.3. IAP element methylation during development	49
1.6.4. IAP element insertions.....	50
1.6.5. Variable methylation within IAP insertions	51
1.7. DNA methylation and the stability of trinucleotide repeats.....	56
1.7.1. Dynamic mutations in human disease	56
1.7.2. CGG trinucleotide repeat expansion at the human <i>FMRI</i> locus.....	59
1.8. Characterization of the imprinted <i>RSVlgmyc</i> transgene.....	67
1.9. Specific aims.....	68
Chapter 2: Methods.....	71
2.1. Generation of transgene constructs	71
2.2. PCR	72
2.3. Generation of transgenic mice	72
2.4. PCR genotyping.....	73
2.5. DNA isolation.....	73
2.6. Southern blot analysis	73
2.7. Collection of preimplantation stage embryos.....	74

2.8. Bisulfite genomic sequencing.....	74
2.8.1. Large DNA samples	74
2.8.2. Preimplantation embryos	76
2.8.3. PCR Primers.....	77
2.9. DNA sequence analysis.....	85
2.10. Sequencing of CGG repeat tracts.....	85
2.11. Blastocyst lambda library preparation.....	85
Chapter 3: Identification of sequences required to create a differentially methylated domain.....	86
3.1. Introduction	86
3.1.1. Mouse model transgenes for the study of genomic imprinting.....	86
3.1.2. Aims of these studies.....	88
3.2. Identification of the <i>RSVlgmyc</i> transgene DMD	89
3.3. Deletion of the <i>RSVlgmyc</i> DMD abolishes imprinting.....	98
3.4. Use of the <i>RSVlgmyc</i> transgene as a model to assess DMD sequence function.....	99
3.4.1. <i>Igf2r</i> DMD sequences functionally replace the <i>RSVlgmyc</i> DMD	100
3.4.2. Not all DMD sequences are able to functionally replace the <i>RSVlgmyc</i> DMD.....	107
3.5. The <i>Igf2rIlgmyc</i> hybrid transgene as a model to study genomic imprinting.....	113
3.6. Analysis of DMD sequences from the paternally methylated <i>H19</i> gene	117
3.7. Design of hybrid transgenes containing maternally methylated DMD sequences	122
3.7.1. Comparison of endogenous gene DMDs.....	122
3.7.2. Tandem repeats restore imprinting to the <i>Ig/myc</i> transgene	123
3.8. Analysis of DMD methylation during development.....	131
3.8.1. The <i>Snrpn</i> DMD sequences function as <i>SnrpnRIlgmyc</i> 's DMD.....	137
3.8.2. DMD methylation is maintained during preimplantation development.....	137
3.8.3. Methylation of endogenous imprinted genes in the blastocyst.....	139
3.9. Discussion.....	146
3.9.1. The mechanisms of maternal and paternal imprinting are distinct	146
3.9.2. Analysis of DMD sequence requirements for maternal imprinting	146
3.9.3. Efficiency of transgene imprinting.....	153
3.9.4. Imprinted gene methylation during preimplantation development.....	154
Chapter 4: Characterization of IAP element methylation in the blastocyst.....	157
4.1. Introduction	157
4.2. Methylation of the general IAP element population.....	158
4.3. Identification of unmethylated IAPs in blastocyst stage embryos.....	161
4.3.1. Experimental design.....	161
4.3.2. Analysis of IAP element methylation.....	165
4.3.3. Comparison of unmethylated IAPs	177
4.4. Discussion.....	180
4.4.1. Single LTRs are unmethylated at the blastocyst stage	180
4.4.2. A specific population of IAPs is unmethylated at the blastocyst stage.....	181
Chapter 5: Analysis of trinucleotide repeat stability using the <i>RSVlgmyc</i> transgene.....	182
5.1. Introduction	182
5.1.1. Mouse models of trinucleotide repeat expansion.....	182
5.1.2. Establishing a transgenic mouse model for trinucleotide repeat expansion	184
5.2. Transgene design	186
5.3. The CGG trinucleotide repeat showed instability following injection	189

5.4. Size changes occurred within the repeat region.....	195
5.4.1. Southern blot analysis.....	195
5.4.2. Direct sequencing of the repeat region.....	204
5.5. The CGG repeat is stable in somatic cells.....	205
5.6. The transgene showed intergeneration stability.....	205
5.7. Maternal methylation of the transgene is not consistent.....	210
5.8. Discussion.....	213
5.8.1. Differences between repeat expansion in the human and mouse.....	213
5.8.2. DNA repair and DNA replication	216
5.8.3. Size changes and transgene injections.....	218
5.8.4. Methylation and repeat stability.....	218
Chapter 6: Summary and Future Directions.....	220
BIBLIOGRAPHY	226

LIST OF TABLES

Table 1. Imprinted loci in the mouse	22
Table 2. Dynamic mutations in human disease	57
Table 3. Bisulfite genomic sequencing PCR primers for endogenous imprinted loci.....	78
Table 4. Bisulfite genomic sequencing PCR primers for transgene loci	80
Table 5. Bisulfite genomic sequencing PCR primers for IAP element LTRs.....	82
Table 6. Summary of <i>pBR/RSV</i> and <i>Ig/myc</i> transgenes	96
Table 7. Summary of hybrid transgenes	110
Table 8. Characteristics of hybrid transgenes for analysis of tandem repeats	129
Table 9. IAP element characteristics	163
Table 10. Description of IAP element methylation and genomic location.....	175
Table 11. CGG repeat expansions and contractions in <i>CGG/Igmyc</i> transgenic lines	200

LIST OF FIGURES

Figure 1. Establishment and maintenance of DNA methylation.....	3
Figure 2. Changes in DNA methylation levels throughout development.....	6
Figure 3. The mammalian DNA methyltransferase proteins.....	9
Figure 4. Establishment and maintenance of DNA methylation at imprinted loci.....	27
Figure 5. DMD methylation at the <i>H19</i> locus.....	33
Figure 6. Maternally methylated imprinted gene clusters in the mouse.....	38
Figure 7. Structure of intracisternal A particles in the mouse genome.....	44
Figure 8. IAP insertions at the <i>agouti</i> locus.....	53
Figure 9. Inheritance of fragile X syndrome and “Sherman’s Paradox”.....	60
Figure 10. CGG repeat expansion at the human <i>FMR1</i> locus.....	63
Figure 11. The imprinted <i>RSVlgmyc</i> transgene.....	69
Figure 12. The DMD of <i>RSVlgmyc</i>	91
Figure 13. Requirement of the <i>RSVlgmyc</i> DMD for transgene imprinting.....	94
Figure 14. Design of hybrid transgenes.....	102
Figure 15. Restoration of transgene imprinting.....	104
Figure 16. Non-imprinted hybrid transgenes.....	108
Figure 17. Loss of Dnmt1o activity affects transgene imprinting.....	114
Figure 18. Paternally methylated DMD sequences do not restore transgene imprinting.....	118
Figure 19. Maternally methylated DMD sequence comparison.....	120
Figure 20. Tandem repeats restore transgene imprinting.....	125
Figure 21. One unit copy of the TR2+3 repeat is not able to restore transgene imprinting.....	127
Figure 22. The bisulfite genomic sequencing technique.....	133
Figure 23. <i>Snrpn</i> DMD sequences are differentially methylated in the transgene.....	135
Figure 24. The <i>SnrpnRlgmyc</i> DMD is differentially methylated during preimplantation.....	141
Figure 25. The endogenous <i>Snrpn</i> gene is differentially methylated in blastocysts.....	143
Figure 26. IAP element LTRs are methylated in sperm and blastocysts.....	159
Figure 27. Identification of unmethylated IAP elements from blastocysts.....	166
Figure 28. Single IAP LTRs are unmethylated at the blastocyst stage.....	170
Figure 29. IAP elements are less methylated in the blastocyst than in the adult.....	172
Figure 30. DNA sequence alignment of unmethylated IAP element LTRs.....	178
Figure 31. Generation of <i>CGG/Igmyc</i> transgenic mice.....	187
Figure 32. Analysis of trinucleotide repeat stability in <i>CGG/Igmyc</i> transgenic lines.....	190
Figure 33. Multiple transgene insertion sites in the <i>CGG/Igmyc-3</i> transgenic line.....	193
Figure 34. The repeat tract changed in size upon generation of founder animals.....	197
Figure 35. Sequencing of the CGG repeat tract contractions.....	202
Figure 36. The CGG repeat is stable in somatic tissues.....	206
Figure 37. The CGG repeat is stable over successive generations.....	208
Figure 38. The <i>CGG/Igmyc</i> transgene is inconsistently differentially methylated.....	211

LIST OF ABBREVIATIONS

bp	basepairs
CpG	cytosine guanine dinucleotide
DMD	differentially methylated domain
DNA	deoxyribonucleic acid
Dnmt	DNA methyltransferase enzyme
ES cells	embryonic stem cells
IAP	intracisternal A particle
kb	kilobases
LTR	long terminal repeat
MBD	methyl-binding domain
mRNA	messenger RNA
nt	nucleotides
PCR	polymerase chain reaction
PGC	primordial germ cell
PWS	Prader-Willi syndrome
RNA	ribonucleic acid
RSV	Rous Sarcoma Virus
SNP	single nucleotide polymorphism
UTR	untranslated region
wt	wild type

Chapter 1: Introduction

1.1. Functions of DNA methylation in mammals

DNA methylation in mammals has several properties that make it essential for distinct processes in the genome. Methylation patterns on genomic DNA are heritable, erasable, and the presence of DNA methylation is associated with transcriptional activation or transcriptional repression (Bestor 2000; Wigler 1981; Prendergast and Ziff 1991). For these reasons DNA methylation is required for the processes of X chromosome inactivation, genomic imprinting, and host defense against mobile genetic elements (Swain et al. 1987; Goto and Monk 1998; Walsh et al. 1998). Each process requires the ability to establish heritable methylation patterns within the genome that are easily reversible and have the ability to control gene expression. X chromosome inactivation silences genes specifically on one X chromosome. Genomic imprinting establishes reversible patterns of DNA methylation in the gametes that are maintained during development and translate into tight transcriptional control. The methylation of mobile genetic elements inhibits their transposition via silencing active transcription, and is an essential mechanism to restrain aberrant insertional mutations in the genome. Methylation is a critical component of each process, and each process is essential for normal mammalian development.

While proper targeting of methylation to specific regions of the genome is essential for normal development, aberrant targeting of sequences for methylation can have adverse effects on development. It is now well established that aberrant methylation of tumor suppressor genes can lead to alterations in gene expression that ultimately result in the development of cancer (Lee 2003). Similarly, at certain disease loci aberrant promoter methylation can lead to silencing of normal gene expression and the development of disease (Pieretti et al. 1991). For example,

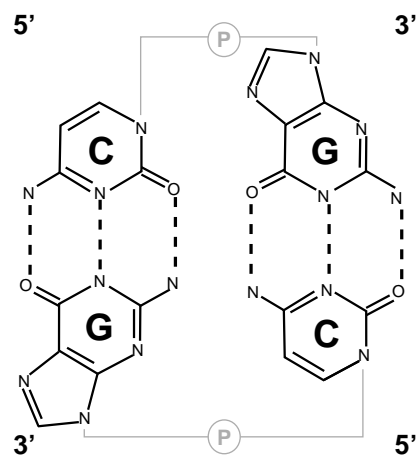
abnormal methylation of the expanded CGG repeat tract at the *FMRI* locus leads to loss of gene expression and development of the fragile X syndrome. For these reasons it is essential to further our understanding of genomic methylation.

1.2. Establishment and maintenance of DNA methylation patterns

DNA methylation in mammals occurs at the 5-position of cytosine in the context of a CpG dinucleotide (Figure 1) (Bestor 2000). Cytosine methylation is established on both strands of the DNA duplex through a process termed *de novo* methylation. A DNA duplex containing symmetrical methylation of the CpG dinucleotide on both strands of the DNA is referred to as fully methylated. Following semi-conservative DNA replication the parent strand of each duplex is methylated and the corresponding daughter strand is unmethylated. This state is referred to as hemimethylated. The pattern of methylation on the parent DNA strand is faithfully copied to the corresponding unmethylated daughter strand through a process termed maintenance methylation (Wigler 1981). DNA methylation can be erased by two possible mechanisms. Passive demethylation of DNA occurs simply by lack of maintenance methylation over successive cell divisions. Active DNA demethylation describes the removal of the methyl group from cytosine by the action of an enzyme termed a demethylase. However, there is currently no known DNA demethylase.

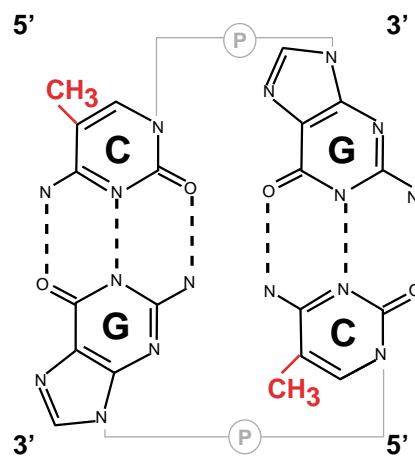
Figure 1. Establishment and maintenance of DNA methylation

DNA methylation is established on an unmethylated DNA duplex (top left) through a process termed *de novo* methylation. *De novo* methylation of DNA in mammals occurs at the 5-position of cytosine in the context of a CpG dinucleotide. The result is a fully methylated DNA duplex on which both cytosines are methylated (top right). Following DNA replication, the parent strand of DNA is methylated and the newly synthesized daughter strand is unmethylated. This is referred to as a hemimethylated CpG dinucleotide (bottom right). The pattern of methylation present on the parent strand of DNA is copied to the daughter strand of DNA through a process termed maintenance methylation. A fully methylated DNA duplex is once again established (bottom left).



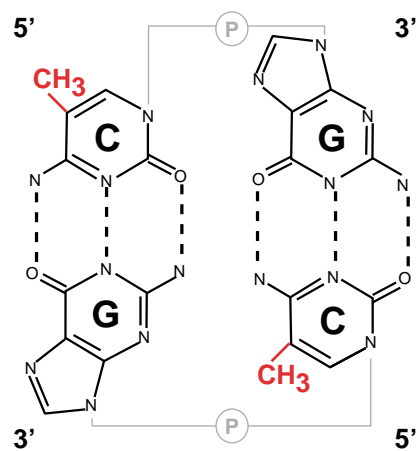
Unmethylated DNA

→
de novo
Methylation



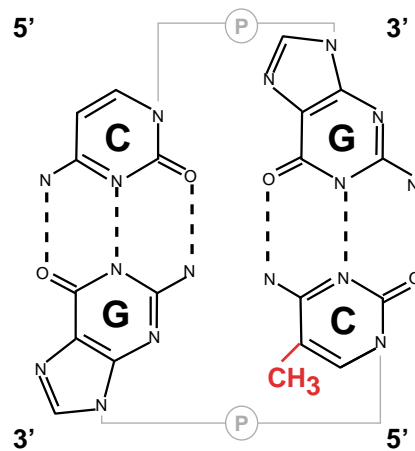
Fully Methylated DNA

↓
DNA Replication



Fully Methylated DNA

←
Maintenance
Methylation

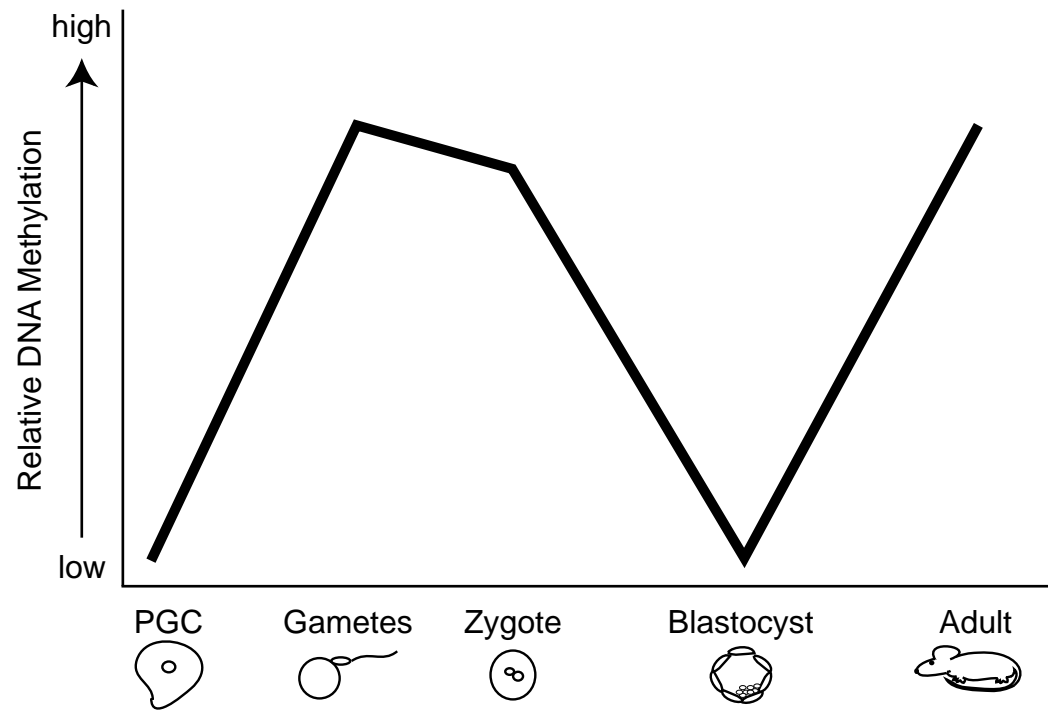


Hemimethylated DNA

Methylation patterns on genomic DNA are dynamic throughout development, exhibiting periods of methylation gain and loss (Figure 2) (Monk et al. 1987; Kafri et al. 1992). Global genomic methylation is erased during the formation of primordial germ cells, the precursors to the gametes. Gamete-specific methylation is then established during spermatogenesis and oogenesis. After fertilization, during preimplantation development, global genomic methylation levels decrease dramatically. Interestingly, the rate of demethylation differs between the maternal and paternal genomes (Oswald et al. 2000). The paternal genome is rapidly demethylated, in a replication-independent fashion, while the maternal genome is slowly demethylated over several cell divisions. The maternal and paternal genomes reach equivalent levels of methylation around the 8-cell stage; however, this timing varies in different mammals. After implantation of the blastocyst stage embryo, high levels of methylation are once again established. The complex patterns of methylation gain and loss are controlled by a group of proteins termed DNA methyltransferases.

Figure 2. Changes in DNA methylation levels throughout development

The graph depicts the increases and decreases in DNA methylation (y-axis) that the genome undergoes throughout various stages of development (x-axis). Methylation is low in primordial germ cells (PGC), the precursors to the gametes. *De novo* methylation occurs during gametogenesis, and high levels of methylation are established in mature sperm and oocytes. Following fertilization and formation of the zygote methylation levels fall dramatically. The majority of the genome is unmethylated by the blastocyst stage of preimplantation development. *De novo* methylation occurs following implantation of the blastocyst stage embryo, and high levels of methylation are once again established. DNA methylation levels are maintained during embryogenesis and in adult somatic cells.



1.3. The mammalian DNA methyltransferases

There are several known mammalian DNA methyltransferase (Dnmt) proteins (Figure 3A). Many of these proteins have been shown to be enzymatically active. Each protein contains a conserved catalytic domain at its C-terminus that is similar to the prokaryotic DNA methyltransferase catalytic domain. The N-terminal domain of each methyltransferase is variable and is thought to be involved in regulating specific protein functions (reviewed by Bestor 2000). The first mammalian DNA methyltransferase to be identified was Dnmt1. The Dnmt2, Dnmt3A, and Dnmt3B proteins were identified by similarity to the catalytic domain of Dnmt1 (Van den Wyngaert et al. 1998; Aapola et al. 2000; Okano et al. 1998). Certain Dnmt proteins are present in multiple isoforms, generated by alternative splicing or use of tissue specific promoters (Weisenberger et al. 2002; Mertineit et al. 1998; Chen et al. 2002). The requirement for many different methyltransferases in mammals suggests that each methyltransferase may have a specific function during development.

1.3.1. DNA methyltransferase 1 activity

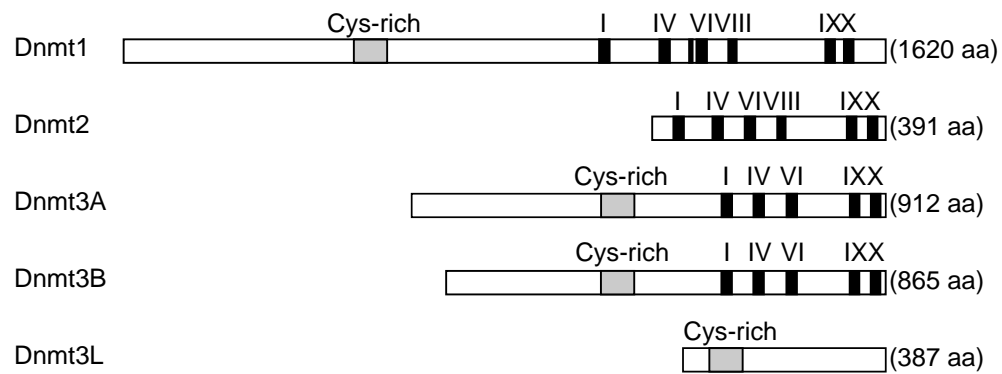
Dnmt1 is the predominant mammalian methyltransferase (Bestor 1988). It has been shown that the Dnmt1 protein is present in at least two isoforms, generated by expression from alternative first exons at the *Dnmt1* locus (Figure 3B) (Mertineit et al. 1998). The *Dnmt1* genomic locus contains three alternative first exons whose expression is driven by tissue specific promoters. The expression of the two Dnmt1 isoforms is tightly regulated at the level of transcription of specific mRNAs from each promoter, and at the level of translation of protein from these mRNAs at specific times during development. The 1 α promoter expresses an mRNA

Figure 3. The mammalian DNA methyltransferase proteins

A, Schematic representation of the mammalian DNA methyltransferase (Dnmt) proteins identified to date. The name of each protein is shown on the far left, and the length of each protein in amino acid number is shown in parentheses on the far right. Black boxes represent conserved motifs within the C-terminal catalytic domain. The C-terminus of the Dnmt3L protein is similar to the Dnmt3A and Dnmt3B proteins; however, it does not contain the conserved motifs required for methyltransferase activity. Each protein, except Dnmt2, contains a conserved cysteine-rich region (gray box) near its amino terminus. B, Schematic of the two isoforms of the Dnmt1 methyltransferase. The somatic isoform (Dnmt1s) is 118 amino acids longer at the amino terminus than the oocyte-specific isoform (Dnmt1o). The length of each isoform in amino acid number is shown in parenthesis to the right. Black boxes represent the catalytic motifs in the C-terminus, and gray boxes indicate regions identified by function, or by homology to other proteins. Figure adapted from figures 2 and 3 (Bestor 2000).

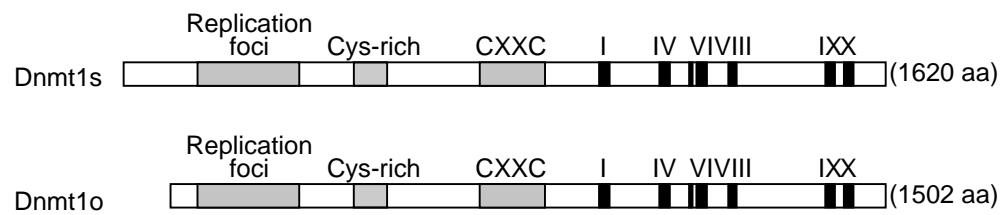
A

Mammalian DNA Methyltransferases



B

Isoforms of the Dnmt1 Methyltransferase



transcript in the oocyte. The 1s promoter expresses an mRNA transcript during preimplantation development and in somatic cells. Another alternative first exon found at the *Dnmt1* locus, designated 1p, is expressed in pachytene spermatocytes and does not encode a protein. Evidence for the presence of other isoforms of the Dnmt1 protein, generated by alternative splicing within the body of the gene, has also been suggested (Hsu et al. 1999; Bonfils et al. 2000; Aguirre-Arteta 2000).

The Dnmt1o and Dnmt1s proteins have very distinct expression and localization patterns. Dnmt1o protein is synthesized in the oocyte, and this protein store is stable during preimplantation development (Mertineit et al. 1998; Howell et al. 2001). Dnmt1o is predominantly localized in the cytoplasm of preimplantation stage embryos except in the growing oocyte and at the 8-cell stage when it is translocated into the nucleus. The Dnmt1o isoform is replaced by the Dnmt1s isoform after implantation of the blastocyst stage embryo. Dnmt1s is then found in the nucleus of all somatic cells in the developing embryo and in the adult. The precise expression and localization patterns of the Dnmt1 isoforms suggest that they have specific functions.

The Dnmt1s and Dnmt1o isoforms are essentially identical, with their only known variation occurring at the amino terminus. The Dnmt1s isoform of the protein contains 118 amino acids at its N-terminus that are not present in the Dnmt1o isoform. The 1s first exon contains an in-frame AUG codon that encodes the 1620 amino acid long Dnmt1s protein. The 1o first exon does not contain an in-frame AUG codon, and translation of the Dnmt1o protein begins in exon 4, generating a protein 1502 amino acids long. Both proteins contain a C-terminal catalytic domain and an N-terminal region that is thought to modulate protein function. Domains for protein-protein interactions, nuclear localization, and targeting to replication foci

have all been identified in the N-terminus (Bestor et al. 1988; Araujo et al. 2001; Fatemi et al. 2001). The Dnmt1o and Dnmt1s proteins have similar enzymatic activities, indicated by the ability of Dnmt1o to functionally substitute for Dnmt1s in somatic cells (Ding et al. 2002). However, the Dnmt1o isoform is 5 times more stable than the Dnmt1s isoform (Ding et al. 2002). This suggests that one primary difference between the two proteins lies in this difference in stability.

Dnmt1 co-localizes with replication foci *in vivo* and has a preference for hemimethylated DNA over unmethylated DNA in *in vitro* (Leonhardt 1992, Bestor and Ingram 1983). This suggests that Dnmt1 travels with the replication fork to maintain methylation following DNA synthesis. It has therefore been proposed that Dnmt1 is primarily a maintenance methyltransferase. Heterozygous mutant mice carrying a hypomorphic *Dnmt1* allele (*Dnmt1^c*) are viable and fertile (Li et al. 1992). Homozygous embryos generated by mating two heterozygous *Dnmt1^{c/+}* animals die by E 9.5 (days post-coitum) of gestation. These homozygous embryos have drastically reduced levels of methylation over the entire genome (Li et al. 1993). The widespread loss of methylation seen with the loss of Dnmt1 activity suggests that this protein is nonspecific, required to maintain methylation everywhere on the genome. ES cells homozygous for the null *Dnmt1^{c/c}* allele still contain a low level of methylation on genomic DNA (Lei et al. 1996). Interestingly, these ES cells still possess the ability to *de novo* methylate newly integrated proviral DNA. At the time these experiments were conducted Dnmt1 was the only known DNA methyltransferase, and these data clearly indicated that other methyltransferases remained to be identified.

1.3.2. Activity of the Dnmt3 family of methyltransferases

The Dnmt3A and Dnmt3B methyltransferases were identified by similarity to the Dnmt1 methyltransferase in its catalytic domain. Both proteins contain the conserved methyltransferase motifs, and are enzymatically active *in vitro* and *in vivo* (Okano, et al. 1998; Hsieh 1999). Dnmt3A and Dnmt3B are expressed at lower levels than Dnmt1 in adult somatic tissues (Okano et al. 1998; Xie et al. 1999). It has been suggested that the Dnmt3A and Dnmt3B methyltransferases are strictly *de novo* methyltransferases. ES cells homozygous for both *Dnmt3a*^{-/-} and *Dnmt3b*^{-/-} null alleles completely lacked the ability to *de novo* methylate newly integrated proviral DNA (Okano et al. 1999). Also, the Dnmt3A protein shows a preference for unmethylated DNA over hemimethylated DNA *in vitro* supporting the notion that it is strictly a *de novo* methyltransferase (Yokochi and Robertson 2002).

Mice heterozygous for individual *Dnmt3a*^{-/+} or *Dnmt3b*^{-/+} mutations are viable (Okano et al. 1999). *Dnmt3a*^{-/-} homozygous mutant mice die at around 4 weeks of age, and *Dnmt3b*^{-/-} homozygous mutant mice die before birth. Mice homozygous for both mutations die during early embryogenesis, and have a more severe phenotype than either individual mutant. These data suggest that Dnmt3A and Dnmt3B have distinct but slightly overlapping functions. Unlike *Dnmt1*^{c/c} mutant embryos, a significant amount of methylation remains on the DNA of embryos lacking both Dnmt3A and Dnmt3B enzymes. Decreases in methylation on specific types of sequences were identified. For example repetitive sequences such as C-type retroviral DNA were undermethylated in *Dnmt3b*^{-/-} embryos, and were even less methylated in *Dnmt3a*^{-/-}, *Dnmt3b*^{-/-} embryos. The loss of both proteins also led to the loss of methylation from the 5' region of the *Xist* gene. Distinct functions of the two methyltransferases are indicated by the specific loss of methylation on minor satellite repeats in the *Dnmt3b*^{-/-} embryos.

Clues to the function of the Dnmt3B protein have come from its human homologue. In humans, mutations in DNMT3B are associated with ICF syndrome (immunodeficiency, centromeric region instability, and facial abnormalities) (Hansen et al. 1999). ICF syndrome is associated with hypomethylation of specific satellite repeats leading to chromosomal instability. This correlates well with the loss of minor satellite repeat methylation specific to *Dnmt3b*^{-/-} embryos. Null mutations in the human DNMT3B protein have never been observed in ICF patients, suggesting that DNMT3B is essential for viability in humans as well as in mice.

The Dnmt3L protein was identified by similarity to the Dnmt3A and Dnmt3B methyltransferases (Aapola et al. 2001). However, Dnmt3L does not possess methyltransferase activity due to amino acid changes in its catalytic domain motifs. Dnmt3L is expressed in oocytes and spermatocytes, and heterozygous *Dnmt3l*^{+/+} mice are viable and fertile (Yoder et al. 2001; Aapola et al. 2001). Homozygous *Dnmt3l*^{-/-} mutant mice are also viable, but males are infertile, suggesting that Dnmt3L is required for spermatogenesis. Dnmt3L co-localizes with Dnmt3A and Dnmt3B in transfected cells (Hata et al. 2002). Additionally, over-expression of Dnmt3L stimulates the activity of Dnmt3A for nonspecific target sequences in a cell culture system (Chedin et al. 2002). These data suggest that Dnmt3A and Dnmt3L interact to regulate *de novo* methylation of DNA target sites.

1.3.3. Possible involvement of Dnmt2 in DNA methylation

Like the Dnmt3 family of methyltransferases the Dnmt2 protein was identified due to its homology with the active Dnmt1 methyltransferase (Yoder and Bestor 1998). Dnmt2 is homologous to the pmt1 protein, a non-essential protein identified in *S. Pombe* (Wilkinson et al. 1995). Fission yeast are not known to methylate their genome and pmt1 shows no *in vitro* activity. Dnmt2 has an intact catalytic domain. However, evidence is equivocal as to whether the

protein possesses methyltransferase activity. It was initially shown that Dnmt2 was not enzymatically active *in vitro* (Van den Wyngaert, et al. 1998). In agreement with these data, *Dnmt2* homozygous mutant embryonic stem cells showed no defects in methylation (Okano et al. 1998). Still, recent studies suggest that Dnmt2 may possess minimal methyltransferase activity (Tang et al. 2003; Hermann et al. 2003; Liu et al. 2003). Therefore, the involvement of Dnmt2 in the process of genomic methylation remains a possibility, and further investigation into this issue is required.

1.3.4. Interactions between the DNA methyltransferase proteins

Studies with the human DNMT1 protein have implicated the methyltransferase in several protein-protein interactions. Human DNMT3A and DNMT3B have been shown to interact with the N-terminus of DNMT1 (Kim et al. 2002). Physical interactions among the DNA methyltransferase proteins would suggest that along with individual functions in establishment and maintenance of genomic methylation, the proteins cooperate to coordinate methylation of the genome. Indeed, functional interactions among these proteins have also been demonstrated (Liang et al. 2002; Fatemi et al. 2002; Rhee et al. 2002). These studies described cooperation among the methyltransferases to maintain methylation in human cancer cells, and in mouse embryonic stem cells. Also, recent studies suggest that the Dnmt3A and Dnmt3L proteins interact to establish methylation at imprinted loci (Hata et al. 2002).

1.3.5. Other proteins required for genomic DNA methylation

Proteins other than methyltransferases are associated with regulation of genomic methylation. The CpG binding protein (CGBP) binds specifically to unmethylated CpGs *in vitro* (Voo et al. 2000; Lee et al. 2001). Targeted mutation of CGBP in the mouse led to homozygous lethality immediately following implantation of the blastocyst stage embryo (Carlone and

Skalnik 2001). Additionally, ES cells generated from homozygous mutant blastocysts show severe defects in genomic methylation (David Skalnik, personal communication). This observation suggests a possible role for CGBP in regulating genomic methylation during early development. The CGBP protein was also shown to act as a methylation-sensitive transcriptional activator, and the possibility exists that CGBP is indirectly involved in regulating methylation during early development (Voo et al. 2000).

ATRX and Lsh, Snf2-like ATP-dependent chromatin remodeling proteins, are both required for maintenance of methylation on genomic DNA (Gibbons et al. 2000; Dennis et al. 2001). ATRX is required for proper methylation of highly repetitive sequences such as rRNA genes, Y specific repeats, and subtelomeric repeats. Lsh is required for methylation of repetitive elements such as IAPs, SINES, LINES, and telomeres. Not unexpectedly, these results demonstrate a requirement for chromatin remodeling proteins in the process of global DNA methylation.

1.4. DNA methylation and the regulation of gene transcription

Many of the functions of DNA methylation require that the methylation mark directly affect the expression of nearby genes. In line with this function, a number of proteins have been described that are able to interact specifically with methylated or unmethylated DNA sequences as a part of their transcriptional regulatory activity. This feature allows the presence of DNA methylation in the promoter region of a gene to affect its ultimate level of expression. These proteins include those that bind specifically to DNA sequences containing methylated CpGs and those that bind specifically to DNA sequences containing unmethylated CpGs. In some cases proteins that interact with the DNA methyltransferases are associated with transcriptional activation or repression.

1.4.1. Methyl-binding domain proteins and transcriptional control

Many proteins have been identified that specifically bind to methylated CpG dinucleotides. These proteins are termed MeCP proteins for methyl-CpG binding protein, and they all contain a conserved DNA binding domain, termed a methyl-binding domain (MBD). The MeCP2 protein was identified biochemically by its ability to bind methylated CpGs, and other members of the family (MBD1-MBD4) were identified by sequence similarity to the methyl-binding domain of MeCP2. Many of these proteins are involved in transcription regulation, either alone or in large transcription regulatory complexes (reviewed by Jorgensen and Bird 2002).

MeCP2 possesses transcriptional repression activity in *in vitro* assays and is a component of the histone deacetylase complex, Sin3a/HDAC (Jones et al. 1998; Nan et al. 1998). Additionally, MeCP2 has a transcription repression domain and the ability to interact with histone methyltransferases, also associated with transcriptional repression (Nan et al. 1997; Fuks et al. 2003). Interactions between methyl-binding domain proteins and histone modifying enzymes may provide a link between DNA methylation and histone modifications to establish transcriptional repression.

Like MeCP2, MBD1 binds methyl-CpGs and contains a transcriptional repression domain (Cross et al. 1997). MBD1 is also involved in repression via association with histone deacetylases and the Suv39h1-HP1 heterochromatic complex (Ng et al. 2000; Fujita et al. 2003). MBD1 contains CXXC domains of unknown function that vary in number in different MBD1 isoforms (produced from different splice variants) (Nakao et al. 2001). The CXXC domain is also found in the Dnmt1 methyltransferase and the CGBP protein and its exact function is unknown.

MBD2 binds methylated CpGs as a part of the MeCP1/NuRD complex that has been shown to remodel and deacetylate chromatin templates (Ng et al. 1999; Feng-Zhang 2001). MBD3, though it contains a methyl-binding domain, is unable to bind methylated CpGs (Wade 2001). However, MBD3 shows an association with the NuRD nucleosome remodeling complex as a co-repressor, and shows a genetic interaction with MBD2 (Zhang et al. 1999). MBD3 is the only MBD protein that is essential for early development (Hendrich et al. 2001).

1.4.2. Methylation-sensitive DNA binding factors

Just as specific proteins are able to bind methylated CpG sequences, binding of certain proteins is inhibited by DNA methylation. Examples of such proteins include YY1, CTCF, SP1, MTF-1, Krox-20 and, c-myc, all of which have effects on gene expression (Prendergast and Ziff 1991; Radtke et al. 1996; Hark et al. 2000; Kim et al. 2003). For example, inhibition of YY1 binding by methylation leads to repression of *Peg3* gene transcription (Kim et al. 2003). As a general trend, methylation of a gene's promoter leads to repression of transcription.

1.4.3. DNA methyltransferase interacting factors

Human DMAP1 was identified by interaction with the human DNMT1 amino terminus (Rountree et al. 2000). The human DMAP1 protein possesses transcriptional repressor activity and also associates with the TSG101 co-repressor. The same study implicated DNMT1 in interactions with the human HDAC2 protein. Similar studies suggest that DNMT1 interacts with E2F1, HDAC1, and the tumor suppressor protein Rb to repress transcription from E2F-responsive promoters (Robertson et al. 2000). Together these results suggest that the methyltransferase may be directly involved in transcriptional repression.

The Dnmt3 family of methyltransferases has also been associated with transcriptional control. In one study the Dnmt3L protein was shown to repress transcription and to co-purify

with HDAC1 (Aapola et al. 2002; Deplus et al. 2002). Both of these activities required the PHD-like motif found within the Dnmt3L protein. Recently, both Dnmt3A and Dnmt3L were shown to interact with SUV39H1 and HP1 α , once again implicating the methyltransferase proteins in the establishment of chromatin (Fuks et al. 2003). All of these interactions increase the types of transcriptional control possible by the establishment and maintenance of a methylated region of DNA.

1.5. Genomic imprinting and DNA methylation

1.5.1. Genomic imprinting

Normal mammalian development requires fertilization of an oocyte by a sperm to form a zygote, with each haploid gamete contributing half of the genetic information to the developing organism. This process suggests that the two parental genomes are equivalent, providing essentially identical information to their offspring. This is the case for most genes; the alleles from each parent contribute equally to the phenotype of their offspring. For example, in humans the allele for brown eye color is dominant to the allele for blue eye color (Eiberg and Mohr 1996). Therefore, a heterozygous child receiving one allele for blue eyes and one allele for brown eyes will have brown eyes, regardless of which parent contributes the allele for brown eyes. This is an example of the simple genetic inheritance described by Mendel. However, many genes have a more complex system of regulation.

Expression of a subset of genes in the mammalian genome is governed by the process of genomic imprinting. Genomic imprinting is defined by the differential expression of a gene depending upon its parental origin. For example, the mouse *Snrpn* gene is expressed when inherited through the paternal germ line, and silent when inherited through the maternal germ line (Leff et al. 1992). Imprinted genes demonstrate consistent parent-specific expression over successive generations, regardless of the sex of the offspring. The *Snrpn* gene is expressed when

passed from a father to his daughter, and the identical allele is silenced when passed from his daughter to his grandson. The opposite is true for the mouse *Igf2r* gene. *Igf2r* is expressed when inherited through the maternal germ line and silenced when inherited through the paternal germ line (Stoger et al. 1993). Due to the parent-specific expression of imprinted genes, the maternal and paternal genomes are not equivalent, even if the DNA sequences of the parental autosomal genes are the same.

1.5.2. Importance of genomic imprinting

Understanding the inheritance of imprinted genes is critical to our understanding of various disease mutations. The occurrence of visible phenotypes from mutations at imprinted loci is altered by the parent-specific expression of the gene involved. Deletion mutations that lead to loss of gene expression are detrimental only when inherited on the normally expressed parental chromosome (Ledbetter et al. 1982). Inheritance of both copies of a chromosomal region from one parent is known as a uniparental disomy (UPD). UPD can have adverse effects due to the loss of expression or overexpression of imprinted genes. For example, the Prader-Willi syndrome results when both copies of human chromosome 15 are maternally inherited, due to loss of expression of paternally expressed imprinted genes found in this region (Nicholls et al. 1989; Glenn et al. 1997). Mutations at imprinted genes are involved in a number of human syndromes including the Prader-Willi syndrome, Angelman's syndrome, and the Beckwith-Weidemann syndrome.

The extent of analysis possible through the study of families with imprinting disorders is limited. Using the mouse model system to study genomic imprinting offers numerous advantages. The process of genomic imprinting is conserved between humans and mice, and many human imprinted genes have mouse homologues that are also imprinted (Leff et al. 1992;

Rachmilewitz et al. 1992). The genetics of inbred mouse strains has been well characterized, and the ability to generate transgenic or knockout animals is a powerful tool. Additionally, the ability to study germ line inheritance, and the short generation time of the mouse makes the system ideal. The large amount of sequence information available for various laboratory mouse strains adds the advantage of using single nucleotide polymorphisms to closely follow the inheritance of imprinted genes. The mouse model has provided an excellent system to study the mechanism of genomic imprinting over the last 20 years.

1.5.3. Genomic imprinting in the mouse

Maternal and paternal genome non-equivalence was first demonstrated in 1984 by pronuclear transfer experiments in the mouse (McGrath and Solter 1984; Surani et al. 1984). Pronuclear transfer between recently fertilized mouse eggs allows the generation of zygotes containing both sets of chromosomes from one parent. Zygotes containing two maternal pronuclei are referred to as gynogenotes, and those containing two paternal pronuclei are referred to as androgenotes. Both gynogenotes and androgenotes were not viable, dying by mid-gestation. Gynogenotes were characterized by fair embryonic tissue development and poor extraembryonic tissue development. Androgenotes were characterized by poor embryonic development and fair extraembryonic tissue development. These experiments clearly illustrated that an embryo must contain information from both parents in order to be viable.

Table 1. Imprinted loci in the mouse

Description of imprinted loci identified in the mouse. Data obtained from <http://www.mgu.har.mrc.ac.uk/imprinting/imprinting.html>. The abbreviations for imprinted loci are listed along with their chromosomal locations and gene names. The repressed parental allele at each locus is also listed. Repressed maternal allele (M). Repressed paternal allele (P).

Imprinted loci	Chromosome	Chromosomal Region	Repressed parental allele; Maternal (M)/ Paternal (P)	Name
Gatm	2	central 2	P	L-arginine: Glucine amidino-transferase
Nnat	2	distal 2	M	neuronatin
Gnas	2	distal 2	P	guanine nucleotide binding protein, alpha stimulating
Gnasxl	2	distal 2	M	guanine nucleotide binding protein, alpha stimulating, extra large
Nesp	2	distal 2	P	neuroendocrine secretory protein
Nespas	2	distal 2	M	neuroendocrine secretory protein antisense
Dlx5	6	centromere to T77H (A3.2)	P	Distal-less homeobox 5
Calcr	6	centromere to T77H (A3.2)	P	Calcitonin receptor
Sgce	6	centromere to T77H (A3.2)	M	sarcoglycan, epsilon
Asb4	6	centromere to T77H (A3.2)	P	Ankyrin repeat and suppressor of cytokine signaling
Peg1/Mest	6	proximal 6 (distal to A3.2)	M	mesoderm specific transcript
Copg2	6	proximal 6 (distal to A3.2)	P	coatomer protein complex subunit gamma 2
Copg2as	6	proximal 6 (distal to A3.2)	M	antisense to Copg2
Mit1/lb9	6	proximal 6 (distal to A3.2)	M	mest linked imprinted transcript 1
Nap1l5	6	proximal 6 (distal to A3.2)	M	Nucleosome assemble protein 1, like 5.
Zim1	7	proximal 7	P	imprinted zinc-finger gene 1
Peg3/Pw1	7	proximal 7	M	paternally expressed gene 3
Usp29	7	proximal 7	M	ubiquitin specific processing protease 29
Zim3	7	proximal 7	P	Zinc Finger Gene 3 from Imprinted domain
Zpf264	7	proximal 7	M	Zinc Finger gene 264
Snrpn	7	central 7	M	small nuclear ribonucleoprotein polypeptide N
Snurf	7	central 7	M	<i>Snrpn</i> upstream reading frame
Pwcr1	7	central 7	M	Prader-Willi chromosome region 1
Magel2	7	central 7	M	Magel2
Ndn	7	central 7	M	necdin
Zfp127/Mkrn3	7	central 7	M	ring zinc-finger encoding gene
Zfp127as/Mkrn3as	7	central 7	M	ring zinc-finger encoding gene antisense
Frat3	7	central 7	M	Frequently rearranged in advanced T-cell lymphomas.
Ipw	7	central 7	M	imprinted in Prader-Willi Syndrome
Atp10c/Atp10a	7	central 7	P	Aminophospholipid translocase
Ube3a	7	central 7	P	E6-AP ubiquitin protein ligase 3A
Ube3aas	7	central 7	M	Ube3a antisense

Imprinted loci	Chromosome	Chromosomal Region	Repressed parental allele; Maternal (M)/ Paternal (P)	Name
Nap114/Nap2	7	central 7	P	
H19	7	distal 7	P	A cDNA clone isolated from a fetal hepatic library
Igf2	7	distal 7	M	insulin-like growth factor type 2
Igf2as	7	distal 7	M	insulin-like growth factor type 2, antisense
Ins2	7	distal 7	M	insulin 2
Mash2	7	distal 7	P	<i>Mus musculus</i> achaete-scute homologue 2
Kcnq1	7	distal 7	P	-
Kcnq1ot1	7	distal 7	M	Kvlqt1 antisense
Tapal/Cd81	7	distal 7	P	Cd 81 antigen
p57KIP2 / Cdkn1c	7	distal 7	P	cyclin-dependent kinase inhibitor 1C
Msuit	7	distal 7	P	mouse specific ubiquitously expressed imprinted transcript 1
Slc2211	7	distal 7	P	Solute carrier family 22 (organic cation transporter member-1 like)
Ipl/Tssc3	7	distal 7	P	imprinted in placenta and liver (Tdag51?)
Tssc4	7	distal 7	P	
Obph1	7	distal 7	P	oxysterol-binding protein 1
A19	9	9	M	
Rasgrf1	9	9	M	Ras protein specific guanine nucleotide-releasing factor 1
Zac1	10	10	M	Zinc finger DNA binding protein
Dcn	10	Distal 10	P	Decorin
Meg1/Grb10	11	proximal 11 (A1-A4)	P	growth factor receptor bound protein 10
U2af1- rs1	11	Proximal 11 (A3.2-4)	M	U 2 small nuclear ribonucleoprotein auxiliary factor related sequence 1
Dlk/Pref1	12	distal 12 (E-F)	M	delta like 1
Meg3/Gtl2	12	distal 12 (E-F)	P	gene trap locus 2
Dio3	12	distal 12 (E-F)	M	Deiodinase Iodothyronine Type 3
Rian	12	distal 12 (E-F)	P	RNA imprinted and accumulated in the nucleus.
Htr2a	14	distal 14	P	5-hydroxytryptamine (serotonin) receptor 2 A
Slc38a4/Ata3	15	distal 15	M	Solute carrier family 38, member 4/Amino acid transport system A3
Peg13	15	distal 15	M	Paternally expressed gene 13
Slc22a2	17	proximal 17	P	Membrane spanning transporter protein
Slc22a3	17	proximal 17	P	Membrane spanning transporter protein
Igf2r	17	proximal 17	P	insulin-like growth factor type 2 receptor
Igf2ras/Air	17	proximal 17	M	insulin-like growth factor type 2 receptor antisense RNA
Impact	18	proximal 18 (A2-B2)	M	Homology with yeast & bacterial protein family YCR59c/yigZ
Ins1	19	19	M	insulin 1

Experiments later showed that inheriting two copies of certain subchromosomal regions from only one parent resulted in abnormal phenotypes in the embryo including embryonic lethality (Cattanach and Kirk 1985). These data further supported the idea that the two parental genomes were not equivalent, and narrowed the focus from entire chromosomes to smaller chromosomal regions. These chromosomal regions were later shown to contain imprinted genes (Kaneko-Ishino 1995; Cattanach et al. 1992; Bartolomei et al. 1991). To date more than 50 genes have been identified in the mouse that are preferentially maternally expressed or preferentially paternally expressed (Table 1). Recent experiments, aimed at identifying novel imprinted genes, indicate that many more imprinted genes remain to be found (Smith et al. 2003; Strichman-Almashanu et al. 2002).

1.5.4. DNA methylation and genomic imprinting

The differential expression of genes based on parental inheritance indicates that both alleles are capable of remembering their parental origin. The reversibility of the process in the gamete suggests that the two alleles are distinguished without altering DNA sequence information. Epigenetic modifications of DNA are defined as heritable changes in gene function that are not due to changes in DNA sequence. These types of modifications can include DNA cytosine methylation, and histone modifications such as acetylation, phosphorylation, methylation, ubiquitination, and ADP-ribosylation. An epigenetic mark capable of distinguishing the alleles of an imprinted gene must be heritable through cell divisions, removable during gametogenesis, and ultimately have an effect on gene expression. DNA methylation makes an excellent candidate for such a mark.

As described above, the cytosine methylation pattern found on genomic DNA is heritable through successive cell divisions. Methylation is established and maintained by a group of

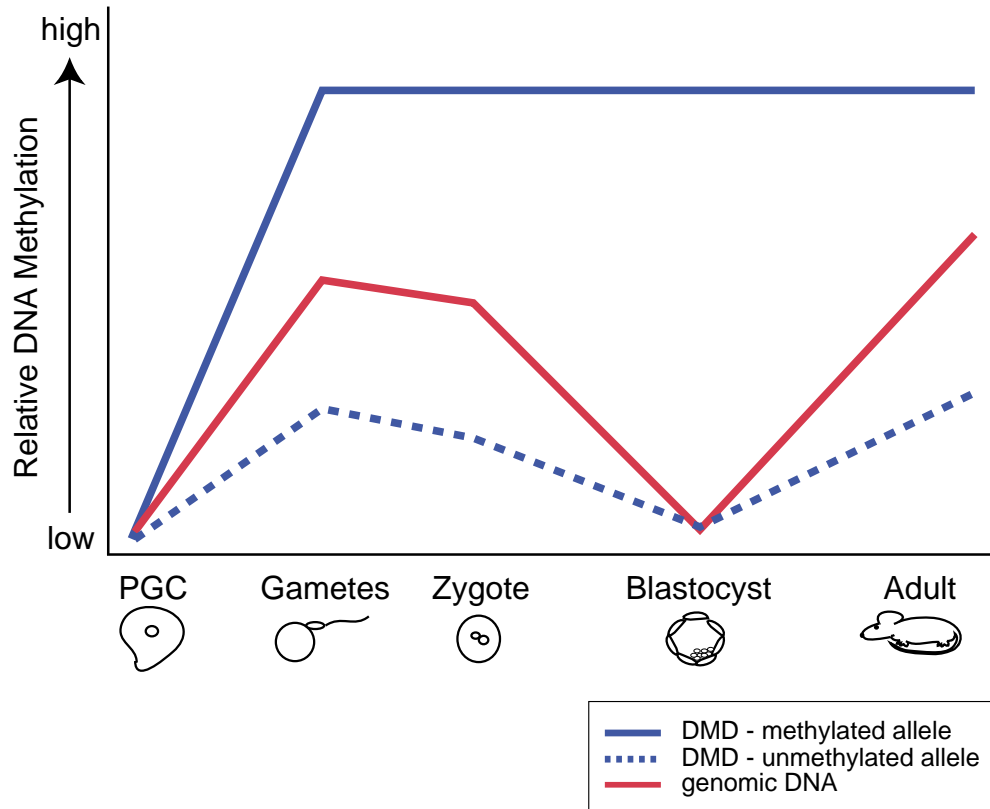
proteins termed DNA methyltransferases (Figure 3A). DNA methylation can also be erased by lack of maintenance methylation over successive cell divisions, or by removal of the methyl group from cytosine by a demethylase. Importantly, DNA methylation has been shown to affect gene expression by inhibiting the binding of transcription factors, or by recruiting methyl-CpG binding proteins (Kim et al. 2003; Hark et al. 2000; Nan and Bird 2001). These characteristics make DNA methylation a suitable epigenetic mark to distinguish the alleles of an imprinted gene.

Many lines of evidence support the prediction that DNA methylation plays an important role in regulating imprinted gene expression. The first genes identified with the differential expression characteristic of imprinted genes were mouse transgenes (Swain et al. 1987; Chaillet 1991). Transgenes have since provided useful tools with which to study the distinguishing features of imprinted genes. The paternally expressed *RSVlgmyc* transgene was shown to contain high levels of CpG methylation on each silent maternal allele, while the active paternal allele was undermethylated (Chaillet 1994). Since, it has been shown that many endogenous imprinted genes contain regions of parent-specific methylation termed differentially methylated domains (DMDs) or differentially methylated regions (DMRs) (Tremblay et al. 1995; Shemer et al. 1997; Stoger et al. 1993). In order for the methylation present on imprinted gene DMDs to be an effective epigenetic mark it must be present at all stages of development, from the gamete to the adult.

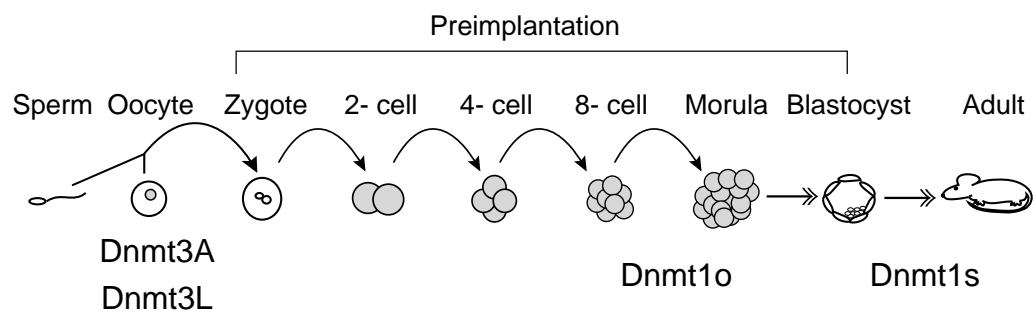
Figure 4. Establishment and maintenance of DNA methylation at imprinted loci

A, The graph depicts the increases and decreases in genomic methylation (y axis) that occur at various stages of development (x axis). Methylation on the bulk of the genome (red line) is compared to the methylated alleles (solid blue line) and the unmethylated alleles (dashed blue line) of imprinted gene differentially methylated domains (DMD). All methylation is erased in primordial germ cells, including the methylated allele of an imprinted gene (PGC). Erasure of methylation in PGCs allows new patterns of gamete-specific methylation to be reset. During gametogenesis methylation differences are established on the methylated and unmethylated alleles of imprinted gene DMDs. During preimplantation development methylation levels decrease throughout the genome, including the unmethylated alleles of imprinted genes. In contrast, methylation on the methylated alleles of imprinted genes is maintained during preimplantation and into the adult. Genomic methylation is established following implantation of the blastocyst stage embryo. However, the unmethylated alleles of imprinted genes remain relatively unmethylated. B, Summary of the stages of development (top) and the methyltransferase proteins that are known to act at specific stages (bottom). Maintenance of methylation at imprinted gene DMDs during preimplantation development requires the activity of methyltransferase proteins. Dnmt3A and Dnmt3L are required to establish maternal-specific methylation at imprinted loci in the oocyte. Dnmt1o is required to maintain maternal- and paternal-specific methylation at imprinted loci in the 8-cell stage embryo. It is not known what methyltransferases are active at the other stages of preimplantation development. Dnmt1s is required for the maintenance of the majority of genomic DNA methylation after implantation of the blastocyst.

A



B



Methylation patterns on global genomic DNA are extremely dynamic during mouse development. Importantly, global genomic methylation levels decrease dramatically after fertilization, and remain low during preimplantation development (Monk et al. 1987; Kafri et al. 1992). In contrast to the methylation gain and loss experienced throughout the genome, imprinted gene DMDs exhibit consistent levels of methylation throughout development (Figure 4A). The maternal and paternal contributions of imprinted genes are physically separated in the gametes, and it is at this time that the differential methylation on imprinted genes is established. It has been demonstrated for some maternally methylated imprinted genes that their DMDs are highly methylated in the mature oocyte and unmethylated in the sperm (Chaillet et al. 1991; Lucifero et al. 2002). It has also been shown that the paternally methylated imprinted gene *H19* is highly methylated in sperm and unmethylated in the oocyte (Lucifero et al. 2002; Tremblay et al. 1995). Interestingly, the gametic methylation present on the paternally methylated *H19* gene and the maternally methylated *RSVlgmyc* transgene is also present during certain stages of preimplantation development. This pattern of methylation is likely to be maintained at all stages of preimplantation development (Chaillet et al. 1994; Tremblay et al. 1997; Warnecke et al. 1998). Specific maintenance of imprinted gene methylation during preimplantation development, a time when most other genomic methylation is lost, suggests that DMD methylation is critical for the perpetuation of an imprint.

The association between DMD methylation and imprinted gene expression is strengthened by the effect of DNA methylation loss on the expression of imprinted genes. A hypomorphic mutation in the *Dnmt1* DNA methyltransferase drastically reduces genomic DNA methylation in homozygous embryos (Li et al. 1992). The loss of methylation is widespread, including the DMDs of imprinted genes. Loss of methylation on imprinted genes in these mutant

embryos abolishes differential gene expression (Li et al. 1993). Together these data suggest that methylation of imprinted gene DMDs is involved in the regulation of imprinted gene expression.

The specificity involved in the acquisition of methylation on imprinted genes in the gamete, and the maintenance of imprinted gene methylation during preimplantation development makes DNA methylation an ideal way to distinguish the alleles of imprinted genes. Unfortunately, little is known about how this specificity is achieved. Presumably the DMD sequences that are targeted for methylation are also involved in directing the process, as well as a number of known and unknown protein factors. Several DNA methyltransferase proteins have clearly been implicated in the regulation of imprinted gene methylation. Other factors have been identified that may be involved in the methylation-dependent expression of imprinted genes. However, it is clear that many factors remain to be identified.

1.5.5. Specific methyltransferases establish and maintain imprinted gene methylation

The Dnmt1s isoform of the Dnmt1 methyltransferase is required for the maintenance of genomic methylation in somatic cells, including the methylation present on imprinted genes (Li et al. 1993). The expression of the Dnmt1o specific isoform during preimplantation development makes it a candidate for a role in maintenance of methylation specific to imprinted genes (Carlson et al. 1992). The *Dnmt1^c* mutation eliminates the Dnmt1s form of the protein in homozygous embryos. However, because Dnmt1o is present as an oocyte store, the function of this isoform cannot be assessed by this method. Therefore, the Dnmt1o specific first exon was targeted for deletion to test its function directly (*Dnmt1^{Δ1o}*) (Howell, et al. 2001). *Dnmt1^{Δ1o/+}* heterozygous mutant mice were viable and fertile, and mating two heterozygous mice generated homozygous wild type, heterozygous mutant, and homozygous mutant mice at expected frequencies. Immunostaining with a Dnmt1 specific antibody showed that oocytes and

preimplantation embryos from homozygous female mice contain no Dnmt1o protein. Interestingly, embryos derived from Dnmt1o-deficient oocytes typically died during the last third of gestation with variable phenotypes and the occasional surviving mouse. This phenotype makes Dnmt1o one of the few maternal effect proteins described in the mouse.

Heterozygous embryos generated from Dnmt1o-deficient oocytes showed no difference in the level of global genomic methylation compared to wild type embryos, however, loss of methylation was seen at imprinted loci. Half of the normally methylated alleles of maternally and paternally imprinted genes were completely unmethylated. This loss of methylation correlated with loss of imprinting. The loss of methylation appeared to be post-zygotic due to the fact that methylation at certain imprinted loci was established normally in Dnmt1o-deficient oocytes. Together these data suggest that Dnmt1o is essential for the maintenance of imprinted gene methylation during preimplantation development. The unique pattern of methylation loss seen in these embryos, and the 8-cell stage nuclear localization of Dnmt1o, suggest that Dnmt1o is only active during the 8-cell stage of preimplantation development. These data also suggest that unknown methyltransferases are still functional at other stages of preimplantation development, resulting in a partial loss of methylation.

Recent experiments suggest that the Dnmt3A methyltransferase is also involved in the process of genomic imprinting. Homozygous *Dnmt3a*^{-/-} mice die at approximately 4 weeks of age, making a detailed analysis of imprinting defects in these mice difficult. However, embryos have been derived from the ovaries of a *Dnmt3a*^{-/-} homozygous mutant female mouse transplanted into a wild type recipient (Hata et al. 2002). These embryos showed a complete loss of methylation at normally maternally methylated imprinted genes. These data suggest that

Dnmt3A is involved in the establishment of methylation on maternally methylated imprinted genes in oocytes.

Although Dnmt3L does not possess methyltransferase activity, many lines of evidence suggest that it also plays a role in controlling imprinted gene methylation. Embryos derived from *Dnmt3L*^{-/-} homozygous oocytes die at approximately E10.5 (Hata et al. 2002). Maternal imprints are not established in these embryos, however paternal imprints are normal. This phenotype is identical to that seen in embryos generated from Dnmt3A-deficient oocytes. This suggests that Dnmt3A and Dnmt3L interact to establish methylation at imprinted genes.

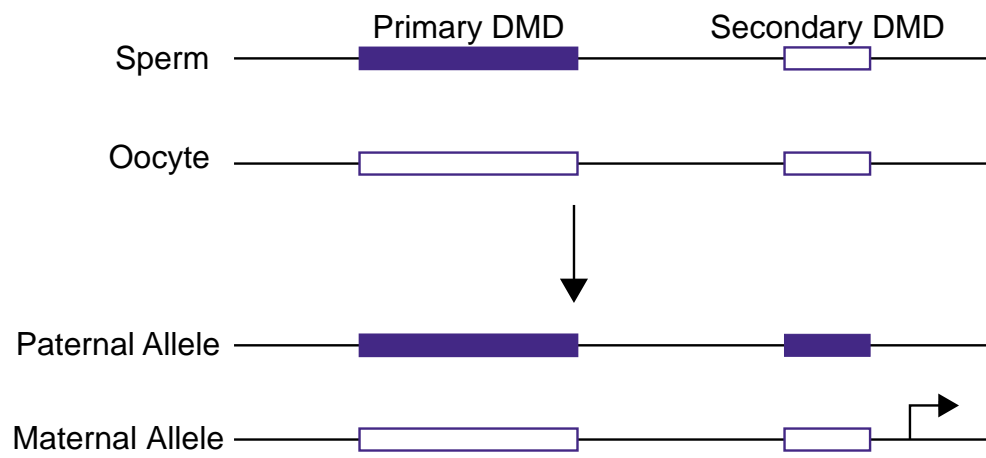
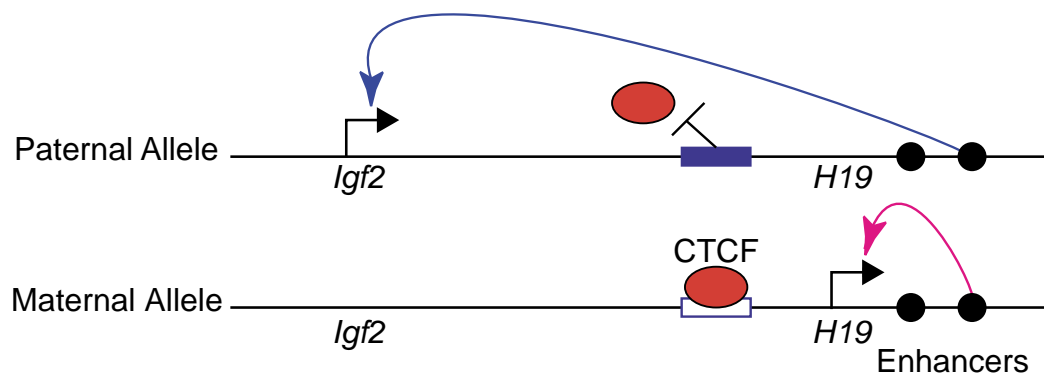
The absolute requirement of DNA methylation to distinguish the parental alleles of imprinted genes is demonstrated by the complete loss of imprinted expression with complete loss of methylation. The need to maintain the identity of each parental allele at all stages of preimplantation development suggests that a methylation dependent step is required at each preimplantation stage (Figure 4B). This model is supported by the unique pattern of methylation loss seen in Dnmt1o-deficient embryos, proposed to miss maintenance methylation specifically at the 8-cell stage (Howell et al. 2001). Moreover, the Dnmt3A and Dnmt3L proteins are specifically required to establish imprinted gene methylation specifically in the oocyte (Hata et al. 2001)). The presence of methylation in the oocyte and at the 8-cell stage suggests that other, unknown methyltransferase proteins are required in sperm and at the other preimplantation stages to coordinate the process.

Figure 5. DMD methylation at the *H19* locus

A, The imprinted *H19* gene has two differentially methylated domains (DMDs) (designated by rectangles). The primary DMD is methylated in the sperm (filled blue rectangle) and unmethylated in the oocyte (open blue rectangle). Paternal-specific methylation of the primary DMD is inherited through all stages of development (bottom). The secondary DMD is equivalently methylated in the gametes, and acquires differential methylation later in development. The paternal-specific methylation at the *H19* locus is associated with maternal-specific expression of *H19* (bent arrow). B, Paternal-specific methylation of the *H19* DMD is required for maternal-specific expression of *H19* and the paternal-specific expression of *Igf2*. Bent arrows indicate active transcription. The methylated DMD is indicated by a filled blue rectangle, and the unmethylated DMD is indicated by an open blue rectangle. Expression of *H19* and *Igf2* requires common enhancer elements located downstream of *H19* (black circles). CTCF (red oval) binds sites in the unmethylated maternal DMD and acts as a boundary element to repress *Igf2* expression (by blocking access to the enhancers), and activate *H19* expression. DMD methylation on the paternal allele inhibits CTCF binding and leads to *Igf2* expression and *H19* repression.

A

Mouse *H19* Gene

**B**

1.5.6. Characteristics of imprinted gene differentially methylated domains

The requirement for gamete specific methylation would suggest that distinct processes operate in oocytes and in sperm to target methylation to the appropriate allele of imprinted genes. A high level of methylation is established on one allele while the other allele is protected from methylation. Subsequently, gamete specific methylation marks are recognized during preimplantation development and maintained during a period of widespread genomic methylation loss. Targeting methylation to the correct sequences requires the action of methyltransferases and other protein factors. It must also require cis-acting sequences in the genome to direct the process. Presumably sequences within the differentially methylated domains themselves are required.

Methylation analyses at many imprinted loci have delineated imprinted gene regions that are differentially methylated in adult tissues (Thorvaldsen et al. 1998; Bressler et al. 2001; Shemer et al. 1997; Tremblay et al. 1997; Yatsuki et al. 2002). Importantly, deletion analyses have shown that at certain loci the regions required for imprinted gene expression overlap those that carry the differential methylation mark. The paternally methylated gene *H19* has made an excellent model for such studies.

A region of differential methylation at an imprinted locus is termed a differentially methylated domain. An essential requirement of a DMD is that it must be established in the gamete and maintained at all stages of development. In order for an imprinted gene to show strict parent-specific expression it is important that differential methylation be established in the gamete, while the parental genomes are separated. Certain DMDs show allele-specific methylation in adult tissues and in the gametes (primary DMDs), while others acquire their methylation post-fertilization (secondary DMDs) (Figure 5A). This suggests that not all

differential methylation found in the adult is necessary to transmit an imprint from the germ line (Hanel and Wevrick 2001; Tremblay et al. 1997).

The CpG methylation patterns at the mouse *H19* locus have been extensively examined. *H19* is paternally methylated and expresses an untranslated RNA specifically from the maternal allele (Tremblay et al. 1997; Bartolomei et al. 1991). Methylation analyses of the *H19* genomic region have shown that it is differentially methylated in the 5' region of the gene, spanning -2 kb to -4 kb from the transcription start site (Figure 5A). Paternal-specific methylation in this region is established in the germ line and is present during many stages of preimplantation development. The promoter region of *H19* is also differentially methylated, however this methylation is acquired after fertilization. These data indicate that the 2 kb region, 5' of *H19*, is its primary DMD.

Targeted deletion studies have demonstrated that *H19*'s DMD is required for maternal-specific *H19* expression. The same region is also required for paternal-specific expression of the *Igf2* gene located 60 kb upstream (Figure 5B) (Thorvaldsen et al. 1998). Both genes are expressed in mesoderm and endoderm tissues and they have been shown to share enhancers located downstream of *H19* (Leighton et al. 1995). Inheritance of a DMD deletion on the normally methylated paternal allele led to paternal activation of *H19* expression and paternal repression of *Igf2* expression. Inheritance of the same deletion on the maternal, unmethylated allele led to repression of maternal *H19* expression and activation of maternal *Igf2* expression. It has recently been shown that the *H19* DMD contains binding sites for the methylation-sensitive boundary element CTCF. CTCF is normally bound at the unmethylated maternal sequences and is necessary for *H19* activation and *Igf2* repression (Hark et al. 2000). CTCF binding is inhibited by methylation of the paternal allele, leading to activation of *Igf2* and repression of *H19*. These

data demonstrate that methylation is important for the maintenance of imprinted *Igf2* and *H19* expression.

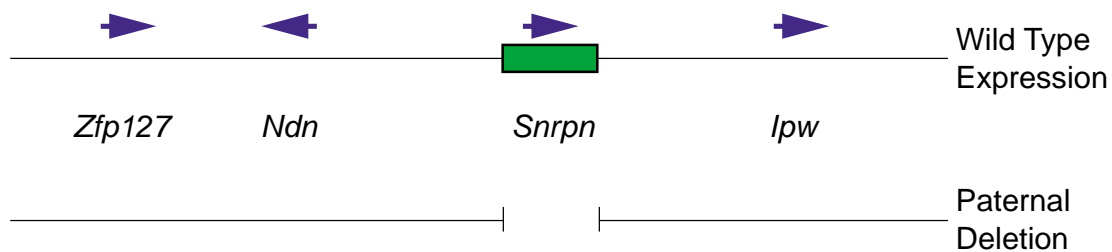
Over fifty imprinted genes have been identified in the mouse, and are distributed on 12 different autosomes (Table 1). Many of these imprinted genes are arranged in clusters of two or more imprinted genes. Similar to the coordinate regulation of *H19* and *Igf2* imprinted gene expression, these genes may be coordinately regulated. Interestingly, genes that do not show parent-specific gene expression are interspersed among the imprinted genes at each cluster. This observation suggests that although coordinate gene regulation in the cluster is widespread it is also very specific within the large chromosomal domain.

Examples of coordinate regulation are found at the mouse *Snrpn*, *Kcnq1*, and *Igf2r* loci (Figure 6). Three paternally expressed genes *Snrpn*, *Zfp127*, and *Ndn* are all located in the same region of chromosome 7. The *Snrpn* imprinted gene contains a maternally methylated DMD that is methylated in the oocyte and in the adult (Shemer et al. 1997). This DMD is involved in the coordinate regulation of each imprinted gene in the cluster. Deletion of the *Snrpn* gene's DMD affected not only *Snrpn* expression, but also affected the expression of other paternally expressed imprinted genes within the *Snrpn* gene cluster (Figure 6A) (Bressler et al. 2001). Inheritance of a DMD deletion through the paternal germ line (normally unmethylated) led to loss of expression of each gene, while inheritance of a DMD deletion through the maternal germ line (normally methylated) had no effect (Figure 6B). This suggests that the unmethylated, paternal DMD is normally involved in gene activation and that methylation or deletion abolishes this activity.

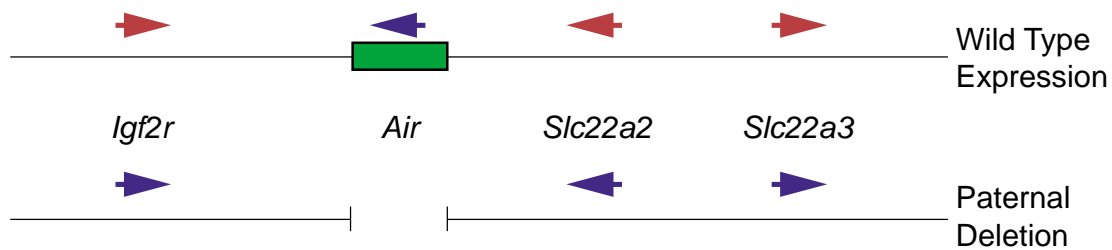
Figure 6. Maternally methylated imprinted gene clusters in the mouse

Diagrams of three maternally methylated imprinted gene clusters in the mouse (not to scale). In each panel the wild type pattern of expression is depicted on the top line. Thin lines represent the genomic DNA, and the location of the DMD in each cluster is shown as a green rectangle. A red arrow indicates maternal-specific expression and a blue arrow indicates paternal-specific expression. The bottom line in each panel shows the effect of a DMD deletion on gene expression on the deleted paternal chromosome. Lack of an arrow indicates a loss of paternal expression. Gain of a blue arrow indicates a gain of paternal expression. At each cluster the pattern of gene expression on the deleted paternal chromosome is identical to that of the methylated maternal chromosome. Biallelically expressed genes are not pictured. A, Expression of imprinted genes in the *Snrpn* gene cluster on chromosome 7. The *Snrpn* gene cluster covers a 2 mb region of chromosome 7 (Gabriel et al. 1998). The DMD of this gene cluster is located within the *Snrpn* gene. B, Expression of imprinted genes in the *Igf2r* gene cluster on chromosome 17. The *Igf2r* gene cluster covers a 400 kb region of chromosome 17 (Sleutels et al. 2002). The DMD of this gene cluster is located within the *Igf2r* gene and contains the promoter for the oppositely imprinted *Air* untranslated RNA. C, Expression of imprinted genes in the *Kcnq1* gene cluster on chromosome 7. The *Kcnq1* gene cluster covers a 600 kb region of chromosome 7 (Fitzpatrick et al. 2002). The DMD of this gene cluster is located within the *Kcnq1* gene and contains the promoter for the oppositely imprinted *Kcnq1ot1* untranslated RNA.

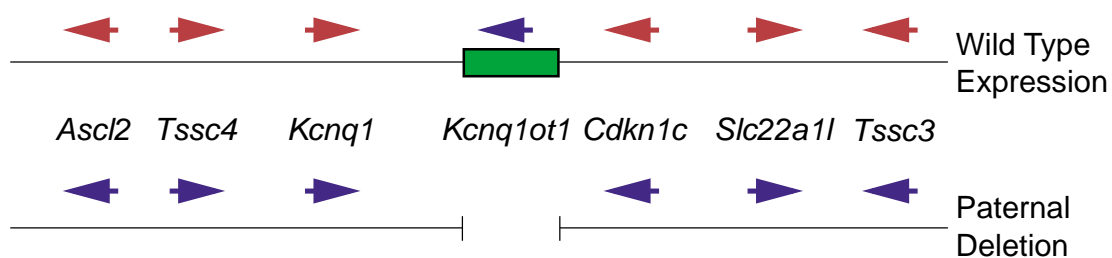
A



B



C



The *Snrpn* gene cluster on mouse chromosome 7 is syntenic to a 2 mb region of human chromosome 15 that contains the human *SNRPN*, *IPW*, and *NDN* genes. The human *SNRPN* gene also contains a maternally methylated DMD (Glenn et al. 1996). Loss of expression of genes in this imprinted gene cluster results in the Prader-Willi syndrome (PWS). Loss of gene expression can be brought about in several ways, including maternal uniparental disomy of chromosome 15, or paternal deletions on chromosome 15. Comparisons of PWS patients with deletions on chromosome 15 have illustrated that a 4.3 kb region is essential for the proper expression of genes within this region (Ohta et al. 1999). The 4.3 kb region deleted in each of these PWS patients includes the promoter and first exon of the *SNRPN* gene, and has been termed the Prader-Willi syndrome-imprinting center (IC). Remarkably, these data suggest that regulation of gene expression in the PWS region is coordinated over a distance of up to 2 mb by the small IC. The PWS-IC overlaps with the DMD located within the *SNRPN* gene. This would suggest that the differential methylation present within this region is involved in the long-range regulation of gene expression. This idea is supported by the fact that certain PWS patients exhibit abnormal methylation at the imprinting center without having identifiable deletions or mutations (Buiting et al. 2003).

The mouse *Igf2r* gene is located within an imprinted gene cluster on chromosome 17 (Figure 6B) (Zwart et al. 2001). Deletion studies at the *Igf2r* locus have shown that its DMD is essential for imprinted expression of multiple genes within the cluster (Wutz et al. 2001). The *Igf2r* gene contains a maternally methylated DMD within its 2nd intron. The DMD houses the promoter for the untranslated *Air* (antisense *Igf2r* RNA) RNA that drives transcription in the opposite direction of *Igf2r* transcription. *Air* shows paternal-specific gene expression, while *Igf2r* and the other genes in the cluster show maternal-specific gene expression (Zwart et al.

2001; Wutz et al. 1997). A DMD deletion on the normally methylated maternal chromosome had no effect on gene expression, while inheritance of a DMD deletion on the normally unmethylated paternal chromosome abolished imprinted expression (Figure 6B). Paternal deletion of the DMD led to loss of *Air* transcription and biallelic expression of the other genes within the cluster.

The *Kcnq1* gene is located within an imprinted gene cluster on chromosome 7, and contains a maternally methylated DMD within intron 10 (Figure 6C). The *Kcnq1* DMD harbors a promoter driving transcription of the paternally expressed *Kcnq1ot1* RNA in the opposite direction of *Kcnq1* transcription (Smilinich et al. 1999). Deletion of the DMD on the normally methylated maternal chromosome had no effect on gene expression, while deletion of the DMD on the normally unmethylated paternal chromosome abolished imprinted expression of multiple genes in the imprinted gene cluster (Fitzpatrick et al. 2002). Paternal inheritance of the DMD deletion led to loss of *Kcnq1ot1* expression and resulted in biallelic expression of the other genes within the cluster (Figure 6C).

The above data point to a requirement for DMD sequences in the appropriate expression of imprinted genes over a long distance. They also suggest that methylation of the sequences is a critical component of the gene regulatory function. Identical patterns of expression are established on the methylated maternal allele and the deleted paternal allele. This suggests that methylation of the maternal allele is functioning to abrogate a gene regulatory activity of the unmethylated DMD sequences.

1.6. Silencing of intracisternal A particles in the mouse genome

1.6.1. Mobile genetic elements

The genomes of both prokaryotic and eukaryotic organisms contain a large number of transposable genetic elements. The presence of transposable elements in the genome was first

suggested by Barbara McClintock in 1951 (reviewed in Barahona 1997). While studying the inheritance of color and the distribution of pigment in maize, McClintock observed that particular genes could be turned on and off at abnormal times, consequently creating variegation in pigmentation. This was suggested to be due to the action of distinct mobile genetic units referred to as "controlling elements" (reviewed by Amariglio and Rechavo 1993). Such genetic elements have since been observed in many organisms including, bacteria (transposons), yeast (Ty elements), *Drosophila* (copia and copia-like elements), and mammals (retrotransposons). In each organism transposition of mobile genetic elements from one location to another in the host genome can cause heritable genetic changes.

Several types of mobile genetic elements have been found in the eukaryotic genome. These elements can be roughly divided into two main groups based on the organization of their genome. The first class includes retrovirus-like elements that contain symmetrical termini, termed long terminal repeats (LTRs). This class of retroelements includes the Ty elements of yeast and the intracisternal A particles (IAPs) of mice. The second class of elements includes those without symmetrical termini, such as short interspersed Alu like repeats (SINEs) and long interspersed repeated elements (LINEs or L1 elements). This second class of elements constitutes the majority of middle repetitive DNA in the mammalian genome. In fact, in humans it has been estimated that approximately 45% of the genome is composed of transposon DNA, while a mere 5% of the genome is made up by cellular genes (Deininger and Batzer 2002). The great abundance of both classes of mobile genetic elements in the genome, in relation to the coding region of the genome, makes them highly significant.

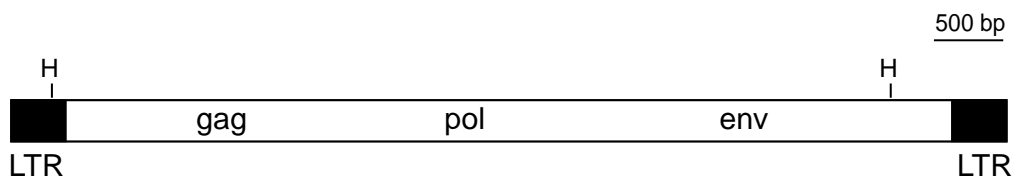
All retrovirus-like elements have a genome organization similar to that of retroviruses such as the HIV virus. The primary distinction between retrovirus-like elements and retroviruses

is that while infective retroviruses are capable of horizontal transmission to other cells, retrovirus-like elements are not able to leave the cell. However, they are capable of replicating their genome and moving to another location within their current host genome, via a mechanism termed retrotransposition. For this reason IAP elements have been termed retrotransposons. Retrotransposition involves transcription of an RNA copy of the retroelement genome by the host RNA polymerase II enzyme. The RNA intermediate is then copied into DNA by a RNA-dependent DNA polymerase encoded by the virus (Review Urnovitz and Murphy 1996). The new DNA copy of the genome can then be integrated into a new location in the host genome. This is a replicative process, the original retrotransposon remains in the same location and a new copy of the retrotransposon is present in another.

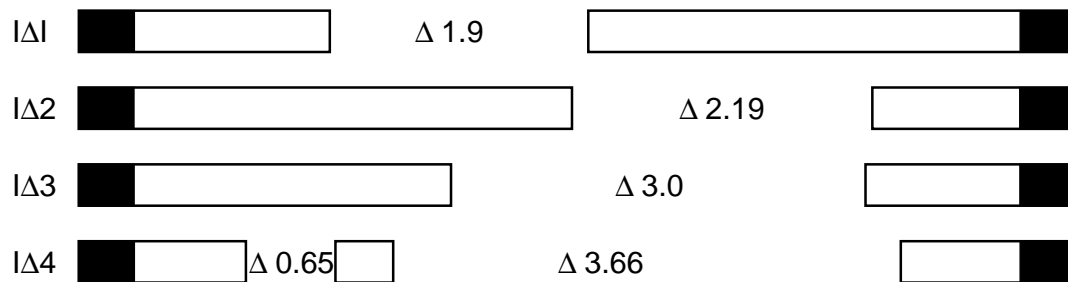
The IAP element is a common type of retrovirus-like element that has been well characterized in the mouse. There are approximately 1000 IAP elements per haploid mouse genome (Review Kuff and Lueders 1988). IAP elements derive their name from the phenotype seen by electron microscopy in cells with actively transcribing IAPs. Retrovirus particles assemble on the golgi apparatus and bud into the cisternae. IAP elements have also been found in the Syrian hamster genome, and it has been suggested that IAP elements may also be found in the human genome (Aota et al. 1987; Garry et al. 1990). IAP elements, in both the mouse and hamster, share a well conserved genome structure (Aota et al. 1987).

Figure 7. Structure of intracisternal A particles in the mouse genome

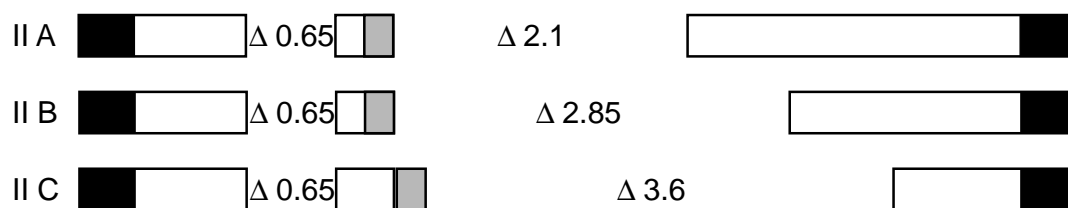
Top, Schematic of a full length type I intracisternal A particle (IAP). The full length type I IAP is approximately 7.2 kb in size and has long terminal repeats (LTRs) on its 5' and 3' ends (black boxes). Conserved *HpaII* sites (H) are indicated above the IAP element. The IAP genome (white rectangle) contains regions similar to the gag and pol regions of functional retroviruses. The env region contains multiple stop codons and is not functional. Bottom, IAP elements are divided into classes based on common internal deletions of the IAP element genome. The locations of deletions are shown by blank spaces within the IAP element genome and deletion sizes are indicated (□ kb). The type II classes of IAP elements all include a 0.5-kb insertion in the gag region (gray box), along with characteristic deletions. Adapted from figure 1 (Kuff and Leuders 1988).



Type I Classes



Type II Classes



The IAP element genome is flanked by LTRs at its 5' and 3' termini (Kuff and Lueders et al. 1988) (Figure 7). The 5' and 3' LTRs of an individual IAP element are essentially identical in sequence. Each LTR is divided into distinct regions termed U3, R, and U5. The U5 region of an IAP LTR is relatively short (50 to 60 bp) and fairly well conserved. The R region, or repetitive region, of an LTR is variable in size, and is composed almost entirely of C and T nucleotides. The U3 region is the most highly conserved region within an LTR and contains the promoter and enhancer elements necessary for transcription of the IAP genome. Transcription of the body of the IAP genome begins at the 3' end of the R region in the 5' LTR and continues until the polyadenylation signal located just following the R region in the 3' LTR. Transcription of the IAP genome is necessary for IAP particle formation and active IAP transposition.

The genome of an IAP element is similar to that of a retrovirus. The body of an IAP element contains three gene regions, termed *gag*, *pol*, and *env* that are found among all retroviruses. The *gag* and *pol* regions of the IAP element are similar to those of the B-type and D-type retroviruses (Fehrmann et al. 2003). The *gag* region encodes a large polyprotein that is cleaved into several structural proteins, and a viral proteinase. These components are necessary for particle formation and budding into the cisternae. The *pol* region encodes a reverse transcriptase and endonuclease, both required for retrotransposition. The *env* region of all known IAP elements contains multiple, conserved stop codons that make this region nonfunctional.

The large number of IAPs in the mouse has been divided into distinct classes, defined by the arrangement of their genome (Figure 7). Type I elements are the major class of IAPs and are approximately 7.2 kb in size when full length. Further subdivisions of the type I elements are

characterized by specific deletions within the IAP gene coding region. Type II elements are classified by an insertion of 500 bp in the *gag* region of the genome. Similar to type I elements, type II elements are further subdivided based on internal genomic deletions. Each IAP element subtype produces a transcript of a unique size. Although many IAP elements contain large gene deletions, IAP elements are still capable of transposition. The largest class of IAP elements are the type I/II elements, which contain a fusion of the *gag* and *pol* gene regions. This class of IAPs is known to be capable of transposition within the mouse genome.

1.6.2. Methylation and control of IAP element transcription

IAP insertions into various genomic locations can affect gene activity in a number of ways. Often IAPs alter transcription levels from the endogenous gene promoter, or interrupt the coding region of a gene. Such IAP element insertional mutations are capable of creating heritable genetic changes that may be deleterious to their host. Unchecked IAP element transposition in the germ line, or even in somatic cells could be detrimental to normal development of the mouse. Therefore, it is thought that the host has developed a defense mechanism to cope with retrotransposon activity. IAP element transposition is dependent upon active IAP element transcription and it has been suggested that transcription from the IAP promoter, and therefore transposition of the IAP genome, is constrained by cytosine methylation of the IAP LTR.

IAP element retrotransposition is dependent upon active transcription of the IAP genome. Therefore restriction of IAP element transcription should be an effective host defense mechanism against unwanted IAP element retrotransposition. IAP element transcription is thought to be inhibited by cytosine methylation within the LTR (Walsh et al. 1998). Many lines of evidence support this hypothesis. It has been shown that *in vitro* methylation of an IAP element LTR is

able to inhibit transcription of a linked reporter construct (Feenstra et al. 1986). Regulation of transcription by methylation is often accomplished by inhibiting the binding of transcription factors to their binding sites. Consistent with this mechanism, the transcription factors EBP-80 and YY1 show methylation sensitive binding to enhancer sequences within the IAP LTR (Falzon and Kuff 1991; Satyamoorthy et al. 1993). Inhibition of enhancer binding activity by methylation alters the level of transcription from the LTR promoter *in vitro*. These data strengthen the connection between IAP element methylation and repression of transcription.

If the ability of methylation to inhibit IAP element transcription *in vitro* correlates with its ability to do so *in vivo*, the level of LTR methylation in various tissues should correlate with the level of IAP element transcription detected in the tissue, and the amount of IAP particles detected in the cell. It has been shown for a number of different tissues and cell lines that the extent of IAP element LTR methylation correlates with the amount of IAP element transcription detected. For example, in normal liver tissue IAP elements are methylated, however, in MOPC-315 myeloma cells IAP sequences are demethylated (Wujcik et al. 1984). These methylation levels correspond to the level of IAP mRNA detected and the number of IAP particles seen in the cells. Moreover, many transformed cells and tumor cell lines show a correlation between IAP LTR demethylation and activation of IAP element transcription (Feenstra et al. 1986; Morgan and Hwang 1984). These data suggest that high levels of methylation are correlated with inactivation of IAP element transcription *in vivo*.

The involvement of DNA methylation in protection of the host genome is also suggested by the effect of loss of global genomic methylation on the transcription of IAPs. Embryos lacking Dnmt1 methyltransferase activity lose methylation on the bulk of their genomic DNA, ultimately resulting in embryonic lethality at E9.5. This loss of methylation is accompanied by a

dramatic increase in IAP element transcription in all regions of E9.5 embryos (Walsh et al. 1998). These data strongly support a role for methylation in controlling IAP element transcription and transposition.

1.6.3. IAP element methylation during development

During the life of the mouse IAP transposition would be most detrimental during gametogenesis and preimplantation development. In order for methylation to make an effective host defense mechanism it should be present on IAP elements at these times in development. Normal levels of genomic methylation are dynamic; methylation is erased in primordial germ cells, established in the gametes, and decrease again after fertilization, falling dramatically by the blastocyst stage (Figure 2). Adult methylation patterns are established after implantation of the blastocyst stage embryo. This pattern of methylation gain and loss is seen over the bulk of the genome; however, there are a few DNA regions that are exceptions to this general trend. One exception is the differentially methylated regions of imprinted genes, which maintain their parent-specific methylation marks throughout preimplantation development (Figure 4). Current data indicate that IAP elements are also an exception to this general trend.

Comparable to the majority of genomic DNA, IAP elements are unmethylated in the primordial germ cells and heavily methylated in the mature oocytes and sperm, limiting the time of demethylation to a few cell divisions. Methylation is also present on the majority of IAP elements at the zygote and preimplantation blastocyst stages (Lane et al. 2003). This indicates that the methylation established on IAP element LTRs in the gametes is maintained throughout preimplantation development. However, recent evidence indicates that the methylation of IAPs at the blastocyst stage is incomplete (Walsh et al. 1998; Lane et al. 2003). Southern blot and bisulfite genomic sequencing data suggest that a portion of the IAP element population is

unmethylated at the blastocyst stage. The incomplete silencing of IAP element transcription during preimplantation development is supported by the fact that IAP element transcripts and IAP particles can be detected in preimplantation embryos. This suggests that some IAP elements are unmethylated and active during preimplantation, or that methylation alone is not fully capable of silencing IAPs.

Maintenance of a silent state for IAP elements is essential even in the somatic tissues. It has been demonstrated that IAP elements are methylated in normal somatic tissues (Mietz and Kuff 1990; Walsh et al. 1998). However, some studies indicate that a small portion of the IAP element population is unmethylated in somatic cells (Mietz and Kuff 1990). Moreover, low levels of IAP transcripts and IAP particles have been detected in some somatic tissues (Mietz et al. 1992; Kuff and Fewell 1985). These data suggest that methylation is a key regulator of IAP transcription, but that some IAP elements escape this inactivation.

1.6.4. IAP element insertions

Many IAP insertion sites have been identified by their effect on gene expression. However, not all IAP insertions are readily detected by easily visible phenotypes. From the C57BL/6J mouse genome sequence it is obvious that many IAP insertion sites are found on every mouse chromosome (UCSC Genome Browser (<http://genome.ucsc.edu/>)). IAP insertions can act in several ways to influence the expression of mRNA from their target genes. An IAP element insertion in the 5' region of a gene can cause ectopic gene expression from the IAP LTR promoter. Alternatively, an IAP insertion 5' of a gene can alter transcription by interfering with endogenous promoter or enhancer elements, or providing enhancer elements from within the IAP LTR itself. An IAP insertion in the coding region of a gene can lead to the production of a truncated or aberrantly spliced mRNA. Along with direct effects, by insertional mutagenesis,

IAP elements can influence gene expression by changing the methylation pattern of the surrounding DNA. Changes in DNA methylation can then lead to changes in neighboring gene expression.

Recent IAP insertions have been found at many loci that have severe effects on normal development. For example, an insertion into the *Ap3d* locus was recently identified in the C3H/HeJ strain of mice (Kantheti et al. 2003). These mice show phenotypes similar to known deletions of the protein, which are a model for human storage pool deficiency (SPD). An IAP element insertion was found 5 basepairs downstream of the start of intron 21. This IAP insertion resulted in expression of an mRNA that was 2 kb larger than the wild type transcript, and produced a C-terminally truncated protein. Interestingly, the IAP insertion in this mouse line is of type I Ψ I, a class of IAP elements commonly associated with active IAP retrotransposition.

Along with disrupting the protein coding region of a gene IAP insertions are able to disrupt normal gene transcription. An example of this type of insertion is seen at the *tyrosinase* (*Tyr*) locus. An IAP insertion at the 5' end of the gene led to decreased gene transcription (Wu et al. 1997). The IAP involved was of type I Ψ I and was inserted in the opposite orientation with respect to the gene. Expression still occurred normally from the *Tyr* promoter, however, it occurred at a lower level. This decrease in expression was attributed to the placement of the IAP between the promoter and necessary enhancer elements located upstream.

1.6.5. Variable methylation within IAP insertions

The most well characterized examples of IAP element insertions are found at the *agouti* locus (*a*). In wild-type mice the pigmentation pattern of the fur is referred to as agouti, meaning a brown-black pigmented hair with a ring of yellow pigmentation just before the tip. Alternating production of eumelanin, brown-black pigments, and pheomelanin, yellow pigments, is normally

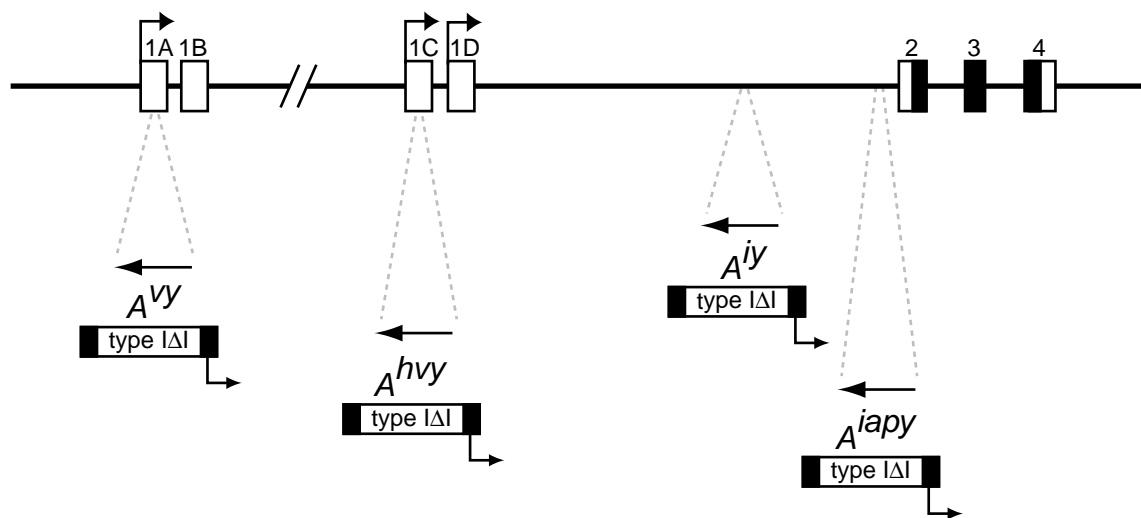
regulated by the *a* locus. The production of agouti protein down-regulates synthesis of eumelanin in favor of the production of pheomelanin. Mutations leading to a high level of expression from the *a* locus result in a mouse with a yellow coat color, and mutations leading to a low level of expression, or no expression, from the *a* locus result in a mouse with agouti coat coloring. Mutations resulting from IAP insertions into the *a* locus are typically dominant mutations, leading to the mis-expression, or overexpression, of the agouti protein and a resultant yellow coat color in carrier animals. Four known insertions, A^{iapy} , A^{vy} , A^{ivy} , and A^{hvy} have similar dominant effects on agouti expression (Figure 8).

The A^{iapy} allele was identified in a cross of a C3H (*A/A*) male to a C57BL/6J (*a/a*) female, which resulted in a spontaneous mutation at the *agouti* locus in a female offspring. This mutation was shown to be due to an IAP insertion located just prior to the first coding exon of the *agouti* locus (exon 2). The inserted IAP element is located in a head to head orientation with the exon (the 5' LTR of the IAP is located adjacent to the 5' end of the exon). IAP promoters are known to be active in both directions. In this case the agouti protein was shown to be produced from a transcript originating in the IAP 5' LTR. Expression from this promoter leads to ectopic expression of the agouti protein in nonexpressing tissues and at abnormal times.

The ectopic expression of agouti protein results in a yellow coat color in carrier animals. Interestingly, all carriers of the A^{iapy} allele did not show a solid yellow coat color, the coat colors ranged from solid yellow to pseudoagouti coat coloring (completely agouti coloring). This range of phenotypes was associated with a corresponding range in expression levels of the *a* mRNA. The expression level of the *agouti* mRNA was shown to correlate with the methylation level of the IAP LTR. Increased methylation led to decreased expression and a pseudoagouti

Figure 8. IAP insertions at the *agouti* locus

Schematic of the *agouti* locus (not drawn to scale) (Miltenberger et al. 2002). Dark lines indicate the genomic DNA and open boxes indicate exons (exon numbers are indicated above each box). Arrows above exons indicate transcription start sites. Translation of the agouti protein begins in exon 2 (shaded boxes). The locations of IAP element insertions are shown below the genomic locus (*agouti* allele names are shown above the IAP). Each IAP insertion is of type I Ψ I. Arrows above each IAP indicate the orientation of the IAP element (pointing 5' to 3'). Promoter activity in the IAP 5' LTR alters expression from the *agouti* locus (bent arrows). (References: Argeson et al. 1996, Michaud et al. 1994, Morgan et al. 1999)



coat coloring. Likewise, decreased methylation led to increased expression and a solid yellow coat coloring. Other variations in coat color arose from intermediate levels of methylation, possibly suggesting mosaicism in the carrier animal. Interestingly, the variation in methylation seen at the *a* locus was parent-specific. Inheritance of the *A^{iapy}* allele from the female parent led to 2.5% pseudoagouti coloring, and inheritance of the *A^{iapy}* allele from the male parent led to 40.6% pseudoagouti coloring. This correlated with a high levels of *agouti* expression in offspring inheriting the *A^{iapy}* allele from the female parent, and a low level of expression in offspring inheriting the *A^{iapy}* allele from the male parent.

Interestingly, the phenotype of the *A^{iapy}* allele is almost identical to the *A^{hy}* allele. The *A^{hy}* allele contains an IAP element inserted in the *agouti* 5' untranslated region in a different location from that of the *A^{iapy}* allele. The only difference in the phenotype generated from these two alleles is the extent of variation seen among the progeny of a specific cross. This similarity suggests that an IAP element present in the context of the *agouti* 5' region is subject to differential regulation of methylation. This supports the notion that the methylation at IAPs is malleable, changing dramatically with the time in development and the location of the IAP.

Changes in DNA methylation are also seen within the vicinity of an IAP insertion near the *Nocturnin* (*mNoc*) promoter (Wang et al. 2001). The IAP insertion present at this locus is found in some inbred mouse strains but absent in others. The IAP insertion is located in the first intron of the *mNoc* gene in DBA/2, BALB/c, C57BL/6J, and C57BL/10 mice. In all tested strains of mice *mNoc* was rhythmically expressed in various tissues, with the pattern of mRNA expression being circadian in nature. The IAP promoter was also active, and produced a hybrid transcript with the *mNoc* open reading frame. Interestingly, the activity of the *mNoc* promoter had an effect on methylation of the IAP element. As the carrier mouse aged IAP transcription

levels increased (Barbot et al. 2002). The increase in mRNA levels was correlated with gradual demethylation of the LTR. This illustrates that surrounding genomic sequence can also exert an effect on the IAP element.

1.7. DNA methylation and the stability of trinucleotide repeats

1.7.1. Dynamic mutations in human disease

A class of mutations, termed dynamic mutations, is occasionally found at human disease loci. Dynamic mutations derive their name from the characteristic large expansion of trinucleotide repeats seen at each locus. At least 14 examples of trinucleotide repeat expansions have been characterized at human disease loci (Reviewed by Cummings and Zoghbi 2000) (Table 2). The Huntington's disease (*HD*) locus contains a CAG trinucleotide repeat within the coding region of the gene. Expansions of this repeat tract lead to expansion of a corresponding polyglutamine tract in the protein. Presence of the expanded polyglutamine tract in Huntington's disease patients results in protein misfolding and abnormal cytoplasmic protein aggregation (Zoghbi and Orr 1999). Trinucleotide repeat tract expansions have also been observed in the 5' UTR (fragile X syndrome), promoter (progressive myoclonus epilepsy type 1), intron (friedreich ataxia), and 3' UTR (myotonic dystrophy) gene regions. Trinucleotide repeats such as CGG, CAG, GAA, and CTG have all been associated with dynamic mutations. The molecular mechanism involved in these trinucleotide repeat expansion disorders is currently not fully understood.

Table 2. Dynamic mutations in human disease

Characteristics of the dynamic mutations known to be associated with human diseases. Data are obtained from Sinden et al. (2002, Review). This is not an all-inclusive list. The disease and associated human locus are listed. For each locus the trinucleotide repeat tract involved is described. The length of the trinucleotide repeat is described for normal individuals (normal), premutation carriers (premutation), or for individuals with the disease (full mutation). Length is shown in number of triplets. Not all trinucleotide repeat expansion disorders are associated with a premutation size range (- or NA), or one may not be known (?).

Disease	Gene	Repeat	Repeat Length		
			Normal	Premutation	Full Mutation
Fragile X syndrome	<i>FMR1</i>	(CGG) _n	30-60	60-200	200-5000
Spinobulbar muscular atrophy (SMA)	<i>AR</i>	(CAG) _n	14-32	?	40-55
Myotonic dystrophy type I (DM1)	<i>DMPK</i>	(CTG) _n	5-37	50-80	80-1000
Huntington disease (HD)	<i>HD</i>	(CAG) _n	10-34	36-39	40-121
Spinocerebellar ataxia 1	<i>SCA1</i>	(CAG) _n	6-44	-	39-82 (pure)
Spinocerebellar ataxia 2	<i>SCA2</i>	(CAG) _n	14-31	-	34-59 (pure)
Spinocerebellar ataxia 3	<i>SCA3</i>	(CAG) _n	13-44	NA	55-84
Spinocerebellar ataxia 6	<i>SCA6</i>	(CAG) _n	4-18	NA	21-33
Spinocerebellar ataxia 7	<i>SCA7</i>	(CAG) _n	4-34	NA	37-306
Spinocerebellar ataxia 12	<i>SCA12</i>	(CAG) _n	7-28	?	66-78
Dentatorubral-pallidoluysian atrophy	<i>DRPLA</i>	(CAG) _n	7-25	?	49-75
Friedreich ataxia (FRDA)	<i>X25</i>	(GAA) _n	6-29	?	200-900

1.7.2. CGG trinucleotide repeat expansion at the human *FMRI* locus

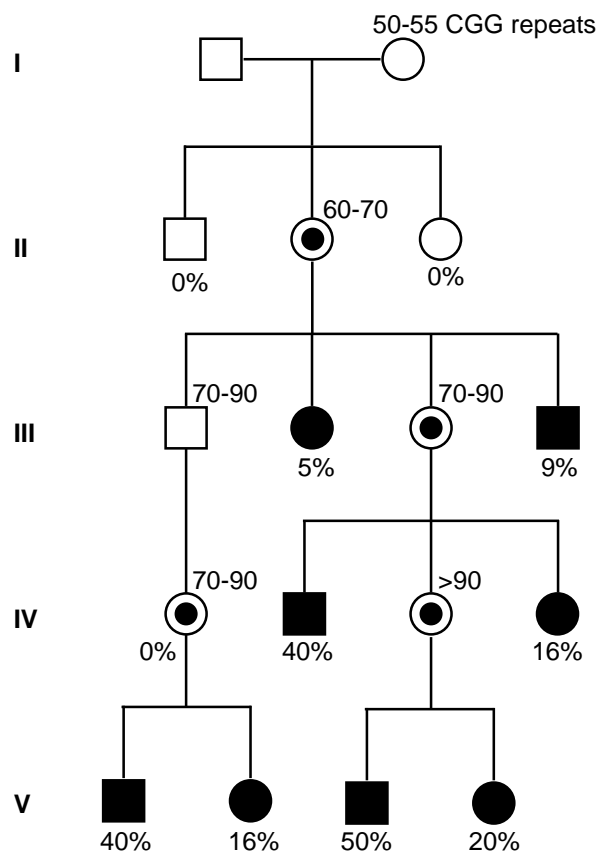
Fragile X syndrome is a common form of mental retardation, affecting approximately 1 in 4000 males and 1 in 8000 females (Jin and Warren 2000 review). Patients with fragile X syndrome are characterized by moderate to severe mental retardation, and have several distinguishing physical characteristics including a large head, long face, prominent ears, and machro-orchidism (Turner et al. 1980; Opitz et al. 1984; Hagerman et al. 1984; Merenstein et al. 1996). Patients are also subject to behavioral abnormalities including poor eye contact, anxiety, and hyperactivity. The mechanism of mutation in fragile X syndrome has been well studied over the years due to its prevalence in the population and the interesting features seen in pedigrees of fragile X families. The inheritance of the disease is similar to that of an ordinary X-linked disease, however, the chance of inheriting the fragile X syndrome varies with an individual's position in the pedigree. For example, a second-generation male had a 9% likelihood of developing the disease, while a fourth-generation male had a 40% chance (Figure 9). This was observed by Sherman in 1985 and has been termed "Sherman's Paradox".

Fragile X syndrome was identified in 1943 by Martin and Bell and was originally termed Martin-Bell syndrome. The disease was later renamed fragile X syndrome when it was found to be strongly associated with a fragile site at the tip of the long arm of the X chromosome (Lubs 1969). Interestingly, the fragile site was shown to correlate with an unstable region of DNA that increased in size following passage through a fragile X pedigree (Yu et al. 1991; Oberle et al. 1991). The fragile site was eventually mapped to the *FMRI* locus (*Fragile X Mental Retardation I*) (Verkerk et al. 1991). The association between the *FMRI* locus and fragile X syndrome was strengthened by the presence of *FMRI* deletion mutations in several fragile X patients (Gedeon

Figure 9. Inheritance of fragile X syndrome and “Sherman’s Paradox”

Inheritance of the fragile X syndrome in a representative family. Symbols: unaffected females (open circles), carrier females (partially filled circles), unaffected males (open squares), fragile X syndrome males (filled squares), and fragile X syndrome females (filled circles). Roman numerals indicate generation number. Percentages located below each symbol indicate the likelihood of acquiring fragile X syndrome. Numbers located above each symbol indicate the number of CGG repeats present at the *FMR1* locus (described in figure 10). The likelihood of an individual acquiring fragile X syndrome increases through successive generations in the pedigree.

(<http://www.ikm.jmu.edu/Buttsjl/ISAT493/Fragile%20X%20Syndrome/fragilexincidence.html>.)



et al. 1992). Eventually, these observations were tied together by experiments demonstrating that the unstable region of the chromosome, and the X chromosome break point, map to a CGG repeat in the 5' untranslated region (UTR) of the *FMRI* locus (Figure 10A). (Kremer et al. 1991; Verkerk et al. 1991). Interestingly, extensive research since this time has shown that the size of the CGG repeat tract is variable among members of the general population, and can expand when inherited through the maternal germ line.

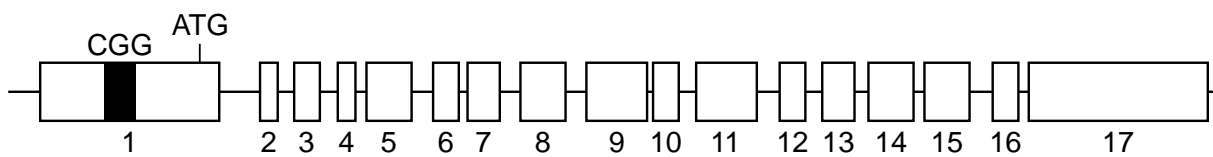
The organization of the CGG repeat tract present at the *FMRI* locus has been well characterized within the population (Figure 10B). The appearance of fragile X syndrome is associated with large expansions of the CGG repeat tract (Fu et al. 1991). The number of CGG repeats within this region is variable, and it has been clearly shown that the number of CGG repeats correlates with the repeat tracts tendency to expand (Yu et al. 1992; Snow et al. 1993). The number of CGG repeats at the *FMRI* locus in normal individuals can range from 30 to 60 CGGs. These relatively short alleles are thought to be stable and have not been associated with expansions. CGG repeat tracts that range from 60 to 200 CGGs have been termed premutation alleles, due to the fact that they are capable of large expansions. Premutation alleles expand to form full mutation alleles, which generally contain 200 to 5000 repeats. Premutation carriers can pass expanded full mutation alleles to their offspring.

How does CGG repeat tract expansion lead to the development of fragile X syndrome? The symptoms of fragile X syndrome are caused by loss of expression from the *FMRI* locus (Pieretti et al. 1991). Initial characterization of the fragile X breakpoint showed that the region of the X chromosome containing the breakpoint was highly methylated in fragile X patients, and unmethylated in unaffected individuals (Vincent et al. 1991; Bell et al. 1991). These data correlate well with data demonstrating that in fragile X patients the CGG repeat tracts are highly

Figure 10. CGG repeat expansion at the human *FMRI* locus

A, Schematic of the human *FMRI* locus (not drawn to scale). Genomic DNA is shown as a black line, exons are depicted as open boxes (exon numbers are indicated below each box). The initiation codon for translation of the FMR1 protein is shown as an ATG. The location of the CGG repeat tract in the 5' untranslated region is indicated by a filled box. B, The CGG repeat region described in panel A is divided into three categories based on its length. Each category is characterized by the number of CGG triplets present at the *FMRI* locus, the presence or absence of expansions, and the presence or absence of methylation at the CGG repeat and associated CpG island. The presence of methylation correlates with an absence of *FMRI* expression.

A



B

Category	Length of Repeat Tract (Number of triplets)	Expansion (Yes (Y); No (N))	Methylation (Yes (Y); No (N))
Normal	30 - 60	N	N
Premutation	60 - 200	Y	N
Full Mutation	200 - 5000	Y	Y

methyated and the *FMRI* gene is silent (Pieretti et al. 1991; Sutcliffe et al. 1992). Methylation at the expanded CGG repeat tract, and a neighboring CpG island, leads to aberrant silencing of *FMRI* transcription, and loss of the FMR1 protein (FMRP) (Pieretti et al. 1991; Sutcliff et al. 1992; Hornstra et al. 1993; Genc et al. 2000). This characteristic expansion and methylation provides an explanation for “Sherman’s Paradox”. Gradual expansion of a premutation allele through the germ line would increase the likelihood of inheriting the fragile X syndrome as you move down in position through the pedigree.

The most interesting aspect of the trinucleotide repeat expansion associated with the *FMRI* locus is its parent-specific mode of inheritance (Malter et al. 1997). The CGG repeat region only expands when a premutation allele is inherited through the female germ line. Because of this, female and male carriers of identical premutation alleles show different frequencies of repeat tract expansions (Nolin et al. 1996). A father carrying a premutation allele will never pass an expanded, full mutation allele to his daughter; however, a mother carrying a premutation allele can pass a full mutation allele to her sons or daughters. Because the *FMRI* gene is located on the X chromosome hemizygous male carriers of a full mutation allele always show symptoms of fragile X syndrome, while heterozygous female carriers can have variable phenotypes due to random X-chromosome inactivation (Abrams et al. 1994). This mode of inheritance suggests that an event specific to the female germ line triggers repeat expansion.

The CGG repeat tract is not completely composed of CGG triplets; it is occasionally interrupted by AGG triplets (Verkerk et al. 1991; Kunst and Warren 1994). The typical sequence of the repeat tract is (CGG)₈₋₁₅ AGG (CGG)₉₋₁₃ AGG (CGG)_x. Periodic AGG interruptions are thought to increase the stability of the repeat region (Gacy et al. 1995). The increase in stability associated with AGG interruptions has a defined polarity. Interruptions at

the 5' end of the repeat tract, that leave a large number of uninterrupted CGG triplets at the 3' end, are not associated with increased stability (Eichler et al. 1996). However, interruptions closer to the 3' end of the repeat tract, that leave a small number of uninterrupted CGG triplets at the 3' end, are associated with increased stability. It is in the 3' region of the repeat tract that large size changes have been documented. The polarity of repeat expansion has led investigators to propose models of expansion that involve errors in DNA replication as the replication fork passes through the repetitive region (Gordenin et al. 1997).

While the predominant size changes seen during inheritance of the CGG repeat tract are expansions, contractions in repeat size have also been documented. Premutation alleles carried in the female germ line have occasionally been shown to revert to the normal size range in some offspring (Brown et al. 1996). Similarly, daughters of premutation carrier males have been shown to contain smaller CGG repeat tracts than their fathers (Fisch et al. 1995). One study, analyzing 191 families, concluded that reverse mutations from father to daughter occurred 22 % of the time (Nolin et al. 1996). These data indicate that the CGG repeat tract is unstable in both directions.

There is considerable debate over the timing of repeat expansion during development. Several factors of repeat expansion point to it being a germ line event. Germ line specific expansion is suggested by the fact that the CGG repeat only expands when inherited from the female parent. In addition, the presence of a single size CGG repeat tract in a fragile X patient suggests that a single expansion event occurs very early in development, either in the germ line or immediately following fertilization (Tassone et al. 1999; Reyniers et al. 1999). However, certain lines of evidence support the notion that CGG repeat expansion also occurs in the mitotic divisions following fertilization. Carriers of premutation alleles and expanded alleles

occasionally show variability in CGG repeat tract length in somatic tissues (Fu et al. 1991; Taylor et al. 1999). Also, monozygotic twins have been shown to have different numbers of CGG repeats at the *FMRI* locus, leading to differences in their expression of the fragile X phenotype (Kruyer et al. 1994). However, mitotic instability of the repeat tract does not exclude the possibility that CGG repeat expansion involves events triggered in the germ line.

1.8. Characterization of the imprinted *RSVlgmyc* transgene

The *RSVlgmyc* transgene was originally designed to express the *c-myc* oncogene in the mouse. It is composed of a variety of sequence elements including pBR322 vector sequences, Rous Sarcoma Virus (RSV) Long Terminal Repeat (LTR) sequences, and a fusion gene from the S107 mouse plasmacytoma cell line (translocation of the *c-myc* gene into the *Immunoglobulin* locus (*Ig*)) (Figure 11A) (Swain et al. 1997). The *c-myc/Ig* sequences within the transgene are organized so that the μ constant and switch regions of the *Ig* locus are immediately followed by the *c-myc* gene with a truncated exon 1. The original *RSVlgmyc* transgene construct was injected into FVB/N fertilized eggs to generate a transgenic mouse line in which over 700 descendents were analyzed. Transgene carriers expressed a transgene-specific *c-myc* transcript only in the myocardium; however, this expression did not lead to an increase in tumor incidence.

Interestingly, expression of the transgene-specific *c-myc* transcript was always observed when *RSVlgmyc* was inherited through the paternal germ line, but never when *RSVlgmyc* was inherited through the maternal germ line. The expressed paternal allele was associated with transgene hypomethylation (a low level of CpG methylation), while the silent maternal allele was associated with transgene hypermethylation (a high level of CpG methylation) in every animal examined. These methylation patterns always correlated with the sex of the transmitting parent and did not vary with the sex of carrier offspring. The methylation patterns remained consistent both in expressing and in non-expressing tissues in each carrier animal, suggesting that they are

the cause of the mono-allelic expression and not a consequence. Differential methylation was subsequently seen in different mouse lines generated with the same transgene construct. Five out of six transgenic lines were maternally hypermethylated and paternally hypomethylated (in one exceptional line ~50% of maternally inherited transgenes showed a paternal pattern of methylation) (Chaillet et al. 1995). These experiments demonstrated that the *RSVImyc* transgene contains all of the sequences necessary to establish an imprint in the maternal germ line, regardless of the site of integration. The consistent imprinting, and easy manipulation of the *RSVImyc* transgene make it an excellent model system to investigate the characteristics of imprinted loci.

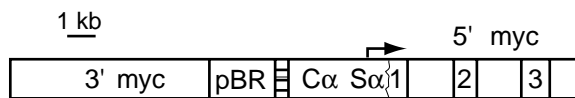
1.9. Specific aims

The preceding introduction described the diverse functions of DNA methylation during mouse development. The aims of this research project focus on gaining a better understanding of different aspects of the establishment and maintenance of DNA methylation during development. The goals of this research project are described in Chapter 3 through Chapter 5. Chapter 3 focuses on the identification of sequences required to create a DMD at an imprinted locus. These experiments used derivatives of the *RSVImyc* imprinted mouse transgene as a model system. Chapter 4 focuses on the characterization of unmethylated IAPs in the mouse genome during the blastocyst stage of preimplantation development. Chapter 5 describes a mouse model system designed to determine the effect of maternal-specific methylation on CGG trinucleotide repeat tract expansion. These experiments used a derivative of the *RSVImyc* imprinted mouse transgene to target methylation to the CGG repeat tract.

Figure 11. The imprinted *RSVlgmyc* transgene

Figure and legend adapted from Chaillet et al. 1995 figure 2. A, Physical characteristics of the transgene. The horizontally hatched area is the 440-bp *PvuII-EcoRI* fragment of the LTR of the Schmidt-Ruppin strain (subgroup D) of RSV. pBR is the *EcoRI-AccI* fragment of pBR322 containing Amp^R. The large *EcoRI* fragment of the circularized construct (bisected by the unique *KpnI* site) contains the breakpoint region of a Burkitt-like immunoglobulin κ /*c-myc* translocation from the S107 plasmacytoma in which the 5' region of the switch recombination sequences of *IgA* (S κ) has been translocated into the 5' region of *c-myc*. Exon 1 of *c-myc* is truncated (and not shown to scale), leaving 23 bp of the 3' end intact (corresponds to bp 968-bp 990 of GenBank locus *Muscmv1*). The arrow represents the approximate start of transcription in S κ of the *IgA/c-myc* fusion transcript. C κ is a 1756-bp *EcoRI-XbaI* fragment (with an internal *XbaI* site) containing intervening sequences and constant region coding exons (corresponds to bp 6374-bp 4620 of GenBank locus *Musialpha*, excluding internal sequence conflicts). Exons 1-3 of mouse *c-myc*. (B) Imprinting characteristics of modified versions of the transgene. For construct *H*, the last 206 bp of C κ has been deleted (corresponds to bp 4825-bp 4620 of *Musialpha*). For construct *I*, the last 355 bp of C κ have been deleted (corresponds to bp 4973-bp 4620 of *Musialpha*). Results are described as a difference or equivalence between maternal and paternal methylation or as expression in the heart. (-) No detectable expression from either maternal or paternal transgene alleles. (M) Maternal; (P) Paternal; (NT) not tested.

A



B

		Methylation	Expression
A	_____	M > P	P
B	_____	M > P	—
C	_____	M > P	P
D	_____	M > P	P
E	_____	M > P	NT
F	_____	M > P	NT
G	_____	M > P	—
H	_____	M = P	NT
I	_____	M = P	NT
J	_____	M = P	NT
K	_____ LacZ	M = P	M = P

Chapter 2: Methods

2.1. Generation of transgene constructs

Hybrid transgenes generated for this study were derivatives of the *RSVlgmyc* transgene (Swain et al. 1997). A number of modifications were made to the original *RSVlgmyc* construct. A unique *NotI* site was created by insertion of a *NotI* linker into a unique *KpnI* site. The transgene was linearized at the *NotI* site and subcloned into the pKS+ plasmid (Stratagene, La Jolla CA); this was done to facilitate removal of pBR322 vector sequences. The transgene was digested with *EcoRI* to remove the pBR322/RSV sequences. This generated a unique *EcoRI* site into which all endogenous mouse sequences were subcloned. Endogenous mouse sequences were amplified by polymerase chain reaction (PCR) from genomic DNA with oligonucleotide primers designed to introduce flanking *EcoRI* sites (described below). Transgenes generated by Mariam Eljanne are described in Reinhart et al. (2002) (*Igf2rIgmyc*, *SnrpnIgmyc*, *H19SIgmyc*, *H19LIgmyc*, and *IAP1gmyc*). The correct sequence and orientation of all transgene constructs was confirmed by sequencing with primers on either side of the inserted sequence. Transgene constructs were prepared for injection by removing pKS+ plasmid sequence by digestion with *NotI*. Digests were run on a 1% agarose gel (Invitrogen, Carlsbad, CA), and the transgene constructs were gel purified using the Qiagen fragment isolation kit. DNA was resuspended in low EDTA, TE (10 mM Tris-HCl pH 7.5, 0.2 mM EDTA pH 8) to a final concentration of 5 ng/ μ l.

PCR primers:

SnrpnRIgmyc transgene: mouse *Snrpn* gene (AF130843, nucleotides [nt] 3796 to 4729)

Forward, GAGAATTCGTGAGCAATCCTTTGG

Reverse, CTGAATTCCTTGAAGCCACATGAAG

Kcnq1lgmyc transgene: mouse *Kcnq1* gene (AF119385, nt 1974 to 2919)

Forward, GGTGAATTCTGGTCCAGTCAGG

Reverse, AAGAATTCCCCTGCTTCTGTAAG

TR2+3Slgmyc transgene: mouse *Igf2r* gene (L06446, nt 1130 to 1307)

Forward, GCTCACGAATTCCCGCGGAACCCTC

Reverse, AGGAGGGAATTCGGAGGGTTTTAGAGGG

2.2. PCR

All sequences used for transgene construction were amplified from C57BL/6J mouse genomic DNA. Each PCR reaction contained: 0.1 μ g of mouse genomic DNA, 1X PCR buffer (Invitrogen, Carlsbad, CA), 1.5 mM MgCl₂ (Invitrogen, Carlsbad, CA), 0.2 mM dNTPs, 0.1 μ M each primer, and 1.25 U of *Taq* DNA polymerase (Invitrogen, Carlsbad, CA) in a 50 μ l reaction mixture. The cycling conditions were as follows: 94° C for 5 min; 30 cycles of 94° C for 45 sec, 60° C for 45 sec, and 72° C for 1 min; and a final extension at 72° C for 10 min. PCR products were resolved on a 1% agarose gel (Invitrogen, Carlsbad, CA) and PCR fragments were isolated using the Qiagen DNA isolation kit. PCR fragments were subcloned into the pCR2.1 vector (TOPO-TA kit, Invitrogen, Carlsbad, CA) and positive clones were sequenced with the M13 reverse primer.

2.3. Generation of transgenic mice

All transgenic mice were created in an inbred FVB/N genetic background by pronuclear injection. Purified DNA fragments were injected at a concentration of 5 ng/ μ l. Injected zygotes were reimplanted into the oviducts of pseudopregnant female Swiss Webster mice and developed to term. Mice were genotyped for integration of the transgene construct by PCR, and Southern

blot. Transgenic founder animals were crossed to wild type FVB/N animals to establish transgenic lines. Transgenic lines were maintained in the inbred FVB/N background.

2.4. PCR genotyping

Mice were genotyped using PCR primers designed to amplify transgene sequence (JRC 94, CTATTCCAGCCTAGTCTGCT and JRC 98, AGTCAGAATCTACGGAGCCT). PCR reactions were performed using approximately 0.1 μ g of mouse genomic tail DNA in a 50 μ l reaction containing, 1X PCR buffer (Invitrogen, Carlsbad, CA), 1.5 mM MgCl₂, 0.1 μ M of each primer, 200 μ M dNTPs, and 1.25 Units of Taq Polymerase (Invitrogen, Carlsbad, CA). The cycling conditions were as follows: 5 minutes at 94°C, followed by 30 cycles of 45" at 94° C, 45" at 58° C, and 45" at 72° C, and a final extension of 10' at 72° C. PCR reactions were resolved on a 1% agarose gel and visualized by ethidium bromide staining.

2.5. DNA isolation

Most DNA samples were isolated by proteinase K (Invitrogen, Carlsbad, CA) digestion (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 10mM EDTA, 0.5% SDS, and 0.2 mg Proteinase K) followed by phenol-chloroform extraction and ethanol precipitation. DNA samples were resuspended in TE (10 mM Tris-HCl pH 7.5, 1mM EDTA pH 8).

2.6. Southern blot analysis

DNA was digested with restriction endonuclease (NEB, Beverly, MA) at the appropriate temperature for 3 hours to overnight. Digests were electrophoresed on an agarose gel (Invitrogen, Carlsbad, CA), and transferred to Genescreen nylon filters (NEN, Boston, MA). Filters were hybridized at 42° C in 40% formamide and with the appropriate ³²P-labeled probe, and washed in 0.1X SSC (1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and 0.1% sodium dodecyl sulfate at 65° C. Bands were visualized by autoradiography. The C α probe is a

1.75 kb *EcoRI-XbaI* fragment of C β , probe a is a 0.6 kb *PstI-BamHI* fragment from the 3' region of *c-myc*, probe b is a 1.3 kb *PvuII-XhoI* fragment of *c-myc* exon 3, and the RSV probe is a 0.44 kb fragment from the LTR region of RSV.

DNA probes for Southern blots were prepared from gel isolated DNA fragments (Qiagen gel isolation kit, Valencia, CA) by random prime labeling. DNA (30 ng) was denatured at 95°C and annealed with random primers (2.5 μ g) on ice. Synthesis of γ -³²P-labeled probe was done in the presence of [γ -³²P] dCTP (50 μ Ci), and Klenow. γ -³²P-labeled probes were purified away from excess unincorporated nucleotides using ProbeQuant G50 micro columns (Amersham Biosciences, Piscataway, NJ).

2.7. Collection of preimplantation stage embryos

All preimplantation stage embryos were collected in M2 media (Specialty Media, Phillipsburg, NJ), washed in M2 medium, washed in 1X phosphate buffered saline pH ~7.3 (PBS) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄•7H₂O, 1.4 mM KH₂PO₄), and stored at -80° C prior to use. Embryo isolation procedures were done as described (Hogan et al. 1994). Fertilized eggs, 2- cell, 4-cell and 8-cell stage embryos were collected from the oviducts at 0.5 days post coitum (dpc), 1.5 dpc, or 2.5 dpc respectively. Blastocyst stage embryos were flushed from the uterus at 3.5 dpc. Extended culture of embryos was done at 37°C and 5% CO₂ in CZB medium (Specialty Media, Phillipsburg, NJ).

2.8. Bisulfite genomic sequencing

2.8.1. Large DNA samples

DNAs for bisulfite genomic sequencing were isolated from tissue samples by grinding in liquid nitrogen prior to proteinase K digestion. DNAs were digested overnight at 37° C with *HindIII*, phenol/chloroform extracted, and ethanol precipitated. DNA was resuspended in TE (10

mM Tris-HCl pH 7.5, 1mM EDTA pH 8) and denatured at a final concentration of 0.3 M NaOH at 42° C for 30 minutes. Denatured DNA was treated with sodium bisulfite (Sigma, St. Louis MO) at a final concentration of 3.06 M and hydroquinone (Sigma, St. Louis MO) at a final concentration of 0.05 mM in the dark at 55° C for 15 to 18 hours. DNA was purified using the geneClean II kit (Qbiogene, Carlsbad, CA) and resuspended in TE. Treated DNA was desulphonated for 15 minutes at 37° C at a final concentration of 0.3 M NaOH, neutralized with 3M NH₄OAc at a final concentration of 0.3 M, and ethanol precipitated. DNA was resuspended in 100 μ l TE and used for PCR analysis. DNA was stored in the dark at -20° C for up to two weeks. Two rounds of PCR were performed. The first reaction contained 1-4 μ l of treated DNA in a 25 μ l reaction containing, 1X PCR buffer (Invitrogen, Carlsbad, CA), 1.5 mM MgCl₂, 0.4 μ M of each primer, 200 μ M dNTPs, and 0.50 Units of Taq Polymerase (Invitrogen, Carlsbad, CA). Cycling conditions for the first round of PCR were: 2 cycles of 94° C for 4 minutes, 55° C for 2 minutes, and 72° C for 2 minutes, followed by 35 cycles of 94° C for 1 minute, 55° C for 2 minutes, and 72° C for 2 minutes, with a final extension of 72° C for 10 minutes. The nested round of PCR was performed with the same reaction conditions and contained 2 - 5 μ l of the first reaction product. Cycling conditions were as follows: 5 minutes at 94° followed by 35 cycles of 94° C for 1 minute, 55° C for 2 minutes, and 72° C for 2 minutes, with a final extension of 72° C for 10 minutes. PCR products were electrophoresed on a 1% agarose gel and isolated using the Qiagen fragment isolation kit (Qiagen, Valencia, CA). PCR fragments were subcloned into the TOPO-TA pCR2.1 vector (Invitrogen, Carlsbad, CA) and sequenced using the M13 reverse primer.

2.8.2. Preimplantation embryos

The bisulfite genomic sequencing technique for preimplantation stage embryos was conducted as described in Schoenherr et al. (2003). Preimplantation stage embryos were stored in 1X PBS at -80°C and thawed at room temperature immediately prior to use. Embryos were briefly centrifuged and embedded in low melting point agarose at a final concentration of 1 to 1.6% [SeaPlaque GTG low melting temperature agarose (Cambrex, Rockland, MD)] Agarose embedded embryos were incubated on ice and covered with cold mineral oil to solidify the bead. DNA was extracted by Proteinase K digestion overnight at 50° C (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS, and 50 µg Proteinase K). The agarose bead was washed three times in 1 ml TE and DNA was denatured with 500 µl of 0.3 M NaOH (two incubations for 15 minutes at room temperature), followed by 0.1 M NaOH (500 µl for 10 minutes at room temperature). The DNA was treated with sodium bisulfite for 5 hours in the dark at 50° C [sodium bisulfite (Sigma, St. Louis, MO) at a final concentration of 3.06 M and hydroquinone (Sigma, St. Louis, MO) at a final concentration of 0.05 mM]. The agarose bead containing treated DNA was washed briefly with water and incubated 5 times for 15 minutes with 1 ml TE. DNA was desulphonated by incubating the agarose bead in 500 µl of 0.2 M NaOH for 15 minutes at room temperature and then at 37°C for 15 minutes, the pH was then neutralized by adding 100 µl of 1 M HCl. The bead was washed briefly in TE, followed by 2 washes for 15 minutes with water. The agarose bead was prepared for PCR by adding water to bring the volume to 50 µl and melting at 65° C for 5 minutes, resuspending the bead and heating to 80° C for 5 minutes. The heated bead was divided into PCR reactions immediately (amount varied for each reaction). Two rounds of PCR were performed. The first PCR reaction conditions were identical to those described for large DNA samples, except the amount of template DNA varied. Prior to the nested PCR reaction the

primary reaction was heated to 80° C for 5 minutes. The nested PCR reaction contained 5 µl of the first reaction product and was performed as described for the large DNA sample. PCR products were electrophoresed on a 1% agarose gel and isolated using the Qiagen fragment isolation kit. PCR fragments were subcloned into the TOPO-TA pCR2.1 vector and sequenced using the M13 reverse primer.

2.8.3. PCR Primers

Primer pairs for endogenous imprinted gene sequences are described in Table 3; primer pairs for transgene constructs are described in Table 4; and primer pairs for IAP elements are described in Table 5. Bisulfite genomic sequencing of endogenous genes was done in embryos obtained from FVB/N mice crossed to B6(CAST7) mice (Mann et al. 2003) and single nucleotide polymorphisms (SNPs) between the two strains of mice were used to distinguish parental alleles. The *H19* region includes a G:A (FVB:Cast) polymorphism at nt position 1566. The *Snprn* promoter region includes a G:T polymorphism at nt position 2348. The *Snprn* repeat region includes a A:G polymorphism at nt position 4097.

Table 3. Bisulfite genomic sequencing PCR primers for endogenous imprinted loci

A and D primers were used for primary reactions and B and C primers were used for secondary reactions. DNA sequences amplified: *H19*, U19616 nt 1301-1732; *Snrpn* promoter, AF081460 nt 2151-2562; *Snrpn* repeats, AF130843 nt 3900-4225.

Endogenous Gene	Primer name	Sequence
<i>H19</i>		
	H19 A	GAGTATTTAAGGAGGTATAAGAATT
	H19 B	GTAAGGAGATTATGTTTATTTTGG
	H19 C	CCTCATAAAACCCATAACTAT
	H19 D	ATCAAAAGTAACATAAAGGGGT
<i>Snrpn</i> promoter		
	Snrpn A	TATGTAATATGATATAGTTTAGAAATTAG
	Snrpn B	AATTTGTGTGATGTTTGTAATTATTTGG
	Snrpn C	ATAAAATACACTTCACTACTAAAATCC
	Snrpn D	AATAAACCCAAATCTAAAATATTTTAATC
<i>Snrpn</i> repeats		
	SnrpnR A	TGGTGGTTTGAGGTTGAGATTGG
	SnrpnR B	GATTTTGGATGTAAGAGTTGTGTTG
	SnrpnR C	ATCCACCCAACCCATAACCCAC
	SnrpnR D	CATCAACAAAACCACAACCTTAAAC

Table 4. Bisulfite genomic sequencing PCR primers for transgene loci

A and D primers were used for primary reactions and B and C primers were used for secondary reactions. The *SnrpnRIgmyc* DMD is over 900 bp in size and was PCR amplified with two separate sets of PCR primers. SnrpnRIgmyc A-1, SnrpnRIgmyc B-1, SnrpnRIgmyc C-2, and SnrpnRIgmyc D-2 PCR primers are specific for the DMD region of the hybrid transgene. These primers were also used to amplify the *TR2+3SIgmyc* transgene DMD. The SnrpnRIgmyc C-1, SnrpnRIgmyc D-1, SnrpnRIgmyc A-2, and SnrpnRIgmyc B-2 are specific for the *SnrpnRIgmyc* DMD.

Transgene	Primer name	Sequence
SnrpnRIgmyc region 1		
	SnrpnRIgmyc A-1	GTATTGAAATTGAGTTTGAAGTGG
	SnrpnRIgmyc B-1	TTGAAGTTATATGAAGTAGTAATAGAG
	SnrpnRIgmyc C-1	CTCTATTAATACAACCAATAAACTAC
	SnrpnRIgmyc D-1	CACTCTAGGATCAAACTATAACTC
SnrpnRIgmyc region 2		
	SnrpnRIgmyc A-2	ATTGGTTGTATTAATAGAGTTATAG
	SnrpnRIgmyc B-2	TTTGGATAATAGAGTGTTTATTTAG
	SnrpnRIgmyc C-2	TATCTTCACCTAAAAACCCTCCAC
	SnrpnRIgmyc D-2	ATACTCTAAATAACCTAAAAAATCC

Table 5. Bisulfite genomic sequencing PCR primers for IAP element LTRs

A and D primers were used for primary reactions and B and C primers were used for secondary reactions. Bisulfite genomic sequencing of IAP 30, IAP 31, IAP 32, IAP 33, IAP 36, IAP 37, IAP 39, and IAP 40 were done using A and B primers specific to each IAP and common IAP C and IAP D primers.

IAP Element	Primer name	Sequence
IAP 23		
	IAP 23 A	TTAGTGTGTGGGTAGATGTTTG
	IAP 23 B	ATGTAAAAAAGTTTGATTAGAGG
	IAP 23 C	TCCCTCTCCAATATTTTAATAAC
	IAP 23 D	CATTTCTCAAATAATATCTTTAC
IAP 7		
	IAP 7 A	GTTTGTTTTATGGGAATTTTTATTA
	IAP 7 B	GTAGTTTTGGTTTTGGAATGAGG
	IAP 7 C	TACCTAACCTTATAATCACATAAT
	IAP 7 D	AATACTACAATATCCAATACATAC
IAP 29		
	IAP 29 A	TTTTTAAGGGTTTGTTATTTTTTTG
	IAP 29 B	AAGGGTTAATTTTTTTGTTTTGTTTA
	IAP 29 C	CAAAAATTAATAACACAATAACAAC
	IAP 29 D	AAAAATATACCAAACCTCAAACC
IAP 35		
	IAP 35 A	GAGTTTTGGGTTATAATATAATGG
	IAP 35 B	TAATTATTAGGTTGAAATGTG
	IAP 35 C	TTTACAACCTTCTTACCATTAAAC
	IAP 35 D	CCTCTTCTCTAAAAAACCTATC
Common		
Common primers used with primers below	IAP C	AACCAAAAAAACACAACAACC
	IAP D	CAATTAAATCCTTCTCAACAATC

IAP Element	Primer name	Sequence
IAP 30		
	IAP 30 A	TGTGGGAGGGTTTAGTTTATTG
	IAP 30 B	GGGAGATGTTATTTTTTGGAGAG
IAP 31		
	IAP 31 A	GAGGGTATAGGGAATTTTTAGG
	IAP 31 B	TAGTATTTGAAATGTAAATAAAG
IAP 32		
	IAP 32 A	TAATATGGAGGTTAGGAAATTGG
	IAP 32 B	TGAATTATGGAGGGGATGATAAG
IAP 33		
	IAP 33 A	GTAAGAGGAGATAGAGGAAGAG
	IAP 33 B	TGGAATAAGATGTAGTAGAATTG
IAP 36		
	IAP 36 A	GTTTATTGGATAGGTAGTTGTTG
	IAP 36 B	ATTAAGTGTAATAGAAAGATTTG
IAP 37		
	IAP 37 A	TTTTTGTGAAGAATGGTTAATTG
	IAP 37 B	AAAATGAGGAAGAGTTGTGTG
IAP 39		
	IAP 39 A	TGTAATAAAGGTTGTTGAGAAG
	IAP 39 B	TTTATTTTTTAAATTGAAATTTAGG

2.9. DNA sequence analysis

All DNA sequencing was performed on the ABI 3700 DNA Analyzer or the ABI 3100 DNA Analyzer (Applied Biosystems, Foster City, CA). Sequences were viewed using Edit View v1.0.1 software (Applied Biosystems, Foster City, CA). All subsequent DNA sequence analysis was performed using MacVector 6.5 software (Oxford Molecular, Atlanta, GA). Pustell DNA matrix comparison was used to find repeat sequences (window size of 20, minimum score 65%). AssemblyLIGN Software was used for DNA sequence alignments (Oxford Molecular, Atlanta, GA).

2.10. Sequencing of CGG repeat tracts

PCR products for sequencing were purified away from primers with the Qiaquick PCR purification kit (Qiagen, Valencia, CA). Template PCR products were denatured for 5 minutes with NaOH at a final concentration of 180 mM NaOH and 9 μ M EDTA. Denatured DNA was precipitated with 40 μ g mussel glycogen as a carrier and washed with 95% ethanol. Sequencing reactions were performed using the Sequenase 7-deaza-dGTP Sequencing Kit (USB, Cleveland, OH). Reactions were run on a 6% denaturing acrylamide gel in 1X TBE (890 mM Tris base, 20 mM EDTA, 890 mM boric acid).

2.11. Blastocyst lambda library preparation

The IAP library screen was performed using FVB/N blastocyst DNA collected from 30 to 60 blastocysts by proteinase K digestion. DNA was digested with *Hpa*II, ligated to adaptor oligos containing *Eco*RI sites, and PCR amplified with an adaptor specific primer and an IAP LTR specific primer containing an *Xho*I site. PCR products were digested with *Eco*RI and *Xho*I and ligated into a predigested ZAP Express vector (Stratagene, La Jolla, CA). Ligation reactions were packaged and plated, and screened as described (Stratagene, La Jolla, CA, Gigapack III Gold Packaging Extract).

Chapter 3: Identification of sequences required to create a differentially methylated domain

3.1. Introduction

3.1.1. Mouse model transgenes for the study of genomic imprinting

As outlined previously, targeted deletion studies in the mouse have clearly illustrated that DMD sequences are required to establish imprinted gene expression. From these studies it is apparent that transcription regulatory sequences are included within the DMD, and that methylation of these sequences on one allele is critical for imprinted expression. However, it is not easy to use deletion analysis to closely define what sequences are required to direct parent-specific methylation in the gamete, and maintain that methylation during development. Large deletions eliminate the target of the methylation, and a fine scale analysis is not easily accomplished by targeted mutagenesis. Because of these limitations, mouse transgenes have been utilized to determine the minimal sequence elements needed to generate an imprint.

Many transgenes have been created that contain sequences from either the *Igf2r* locus or the *H19* locus. Small transgenes composed of the *H19* DMD and 4 kb of genomic flanking sequence were not capable of creating an imprint (Cranston et al. 2001). *H19* transgenes containing 14 kb to 16 kb of genomic sequence, including the DMD, were not consistently imprinted, suggesting that they contained some, but not all of the required imprinting elements (Cranston et al. 2001; Elson and Bartolomei 1997). Only large transgenes comprised of 150 kb of genomic sequence surrounding the *H19* DMD were consistently imprinted (Ainscough et al. 1997). Similar results were obtained from transgenes containing the *Igf2r* DMD. The *Igf2r* DMD alone was not imprinted (Wutz et al. 1997). However, a 130 kb transgene containing the DMD was imprinted. Imprinting of this transgene was abolished when the DMD was removed

(Wutz et al. 1997). These data suggest that the cis-acting sequences required to imprint a transgene are spread over a large distance at the genomic locus. Thus, while transgene studies have confirmed that the DMD is essential for establishing an imprint, the inability to generate small, consistently imprinted transgenes makes these analyses difficult to interpret.

While transgene studies using endogenous gene sequences have been challenging and often hard to interpret, a number of experiments using transgene model systems have contributed valuable information to the field of imprinting. Work over the last 15 years using the *RSVlgmyc* transgene has provided information critical to our understanding of the epigenetic regulation of genomic imprinting. *RSVlgmyc* was one of the first genes identified that exhibited the characteristics of an imprinted locus (Swain et al. 1987). *RSVlgmyc* expresses the *c-myc* gene exclusively from the paternal allele. Correspondingly, *RSVlgmyc* is always highly methylated when inherited through the maternal germ line, and always undermethylated when inherited through the paternal germ line (Chaillet et al. 1995). Parent-specific expression and methylation of the transgene are observed at all sites of integration (Chaillet et al. 1995). The consistent imprinting of the transgene illustrates that its imprinting is controlled by sequence elements contained within the transgene itself, and does not rely on flanking genomic sequence. In essence, the transgene is a portable imprinted locus.

The above data suggest that removal of specific sequence elements from the transgene would abolish imprinting. However, the results obtained from transgenic lines carrying deletion constructs did not provide a clear answer (Figure 11B) (Chaillet et al. 1995). Transgenes identical to *RSVlgmyc*, but lacking the pBR322 sequences, were still maternally methylated and paternally expressed (construct C). Similarly, transgenes carrying a deletion of the RSV LTR sequences showed maternal specific methylation (construct B). The RSV LTR contains

necessary enhancer sequences and transgenes lacking these sequences were not expressed. Deletion of the C \square region alone also had no effect on transgene methylation; maternal alleles were highly methylated and paternal alleles were undermethylated (construct *D*). Likewise, deletion of *c-myc* and S \square together did not have an effect on transgene methylation (construct *F*) (removal of the S \square region eliminates a cryptic promoter and completely eliminates transgene expression). Only when a 206 base pair region at the junction of the C \square and S \square regions was removed from a minimal, imprinted transgene construct was an effect on transgene imprinting seen (construct *H* compared to construct *F*). Equivalent, high levels of methylation were found on both alleles of this transgene. Smaller transgene constructs, also eliminating this sequence, were not imprinted (constructs *I* and *J*).

The above series of experiments would suggest that the 206 base pair region is essential for establishing an imprint. However, a transgene containing this 206 bp sequence and its immediate surrounding sequence, fused to a *LacZ* reporter construct, was not imprinted (construct *K*). Moreover, the *Ig* locus itself is not differentially methylated or expressed. The local environment of the 206 base pair sequence must also affect differential transgene methylation. In summary, it seems that the unique combination of sequences within the *RSVlgmyc* transgene is required to establish an imprint, and not one easily definable element.

3.1.2. Aims of these studies

Little is known about what attracts methylation to DMD sequences in just one germ line, or how this parent specific methylation is maintained. It is assumed that the differential methylation present at imprinted gene DMDs in the germ line is maintained at all stages of development; however, imprinted gene methylation has not been analyzed at all stages of development, for any individual imprinted gene. We would therefore like to develop a model

system to identify sequences in a DMD that are required to establish and maintain parent-specific methylation. We would also like to determine if the methylation on these sequences is maintained during each stage of preimplantation development. In order to attain these goals we have used the *RSVImyc* imprinted mouse transgene as a basis to develop a model system.

The first goal of this set of experiments was to define the differentially methylated domain of *RSVImyc* (Reinhart et al. 2002). This was done by comparing the methylation of CpG dinucleotides within different regions of the transgene after either maternal or paternal inheritance. The requirement of this DMD for transgene imprinting was demonstrated by Mariam Eljanne. The non-imprinted transgene (lacking the DMD) was then used to identify sequences within endogenous imprinted gene DMDs that are required to establish a differentially methylated domain. Lastly, this transgene system was used to analyze the methylation of imprinted gene DMD sequences during preimplantation development.

3.2. Identification of the *RSVImyc* transgene DMD

Methylation of the transgene has previously been characterized by Southern blot analysis using a variety of probes and methylation sensitive restriction endonucleases. Methylation sensitive restriction endonucleases are unable to cleave methylated DNA, but are able to cleave unmethylated DNA. In this way, the methylation of specific CpG sites can be assessed. Southern blot analysis was performed using the methylation sensitive restriction endonuclease *HpaII* (recognition site CCGG) to compare the methylation of the maternal and paternal transgene alleles. DNAs were collected from hemizygous carriers of the transgene after maternal or paternal inheritance, digested with *HpaII* and hybridized with a probe to the C α region of the transgene (Chaillet et al. 1995). The *RSVImyc* transgene has a high level of methylation after maternal inheritance. A high level of methylation is indicated by an intense, high molecular

weight band of uncut DNA. In contrast, the transgene has a low level of methylation after paternal inheritance, indicated by a pattern of low molecular weight bands.

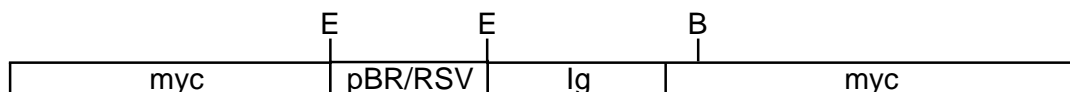
One key feature of many known endogenous imprinted genes, and the *RSVlgmyc* transgene, is their ability to establish and maintain maternal-specific methylation throughout development. In order to determine if the *RSVlgmyc* transgene contains a region homologous to an endogenous imprinted gene DMD, the methylation at various regions of the transgene was examined by Southern blot analysis. DNA samples were cleaved with methylation insensitive restriction endonucleases to target specific regions of the transgene and methylation sensitive restriction endonucleases to analyze CpG methylation. Southern blots were hybridized with DNA probes corresponding to the region of interest.

Southern blot analyses were performed on the *RSVlgmyc* DNA samples described above. The pBR322 and RSV sequences of the transgene comprise a 2.5 kb region (pBR/RSV) that is flanked by *EcoRI* sites (Figure 12A). To examine the methylation in the pBR/RSV region of the transgene, DNAs were first digested with *EcoRI*, and then digested with the methylation sensitive enzymes *HpaII*, *HhaI*, or *BstUI* (recognition sequences CCGG, GCGC, and CGCG, respectively) (Figure 12B) (Reinhart et al. 2002). Resulting DNAs were hybridized with the RSV probe. Following maternal inheritance of *RSVlgmyc* the pBR322/RSV sequences were methylated at each CpG site tested. In contrast, the same region of the transgene was unmethylated following paternal inheritance. These data clearly illustrate that the pBR/RSV region of the transgene is differentially methylated.

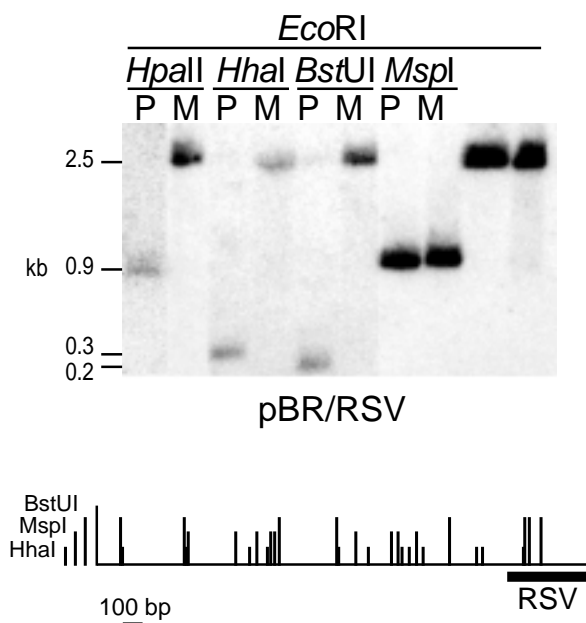
Figure 12. The DMD of *RSVlgmyc*

A, Schematic of the *RSVlgmyc* transgene linearized at the unique *KpnI* restriction site. The *myc* designates exons 1, 2, and 3 of the *c-myc* endogenous locus and the 3' noncoding genomic sequence of *c-myc*. *pBR/RSV* refers to *pBR322* vector sequences and RSV long terminal repeat sequences. *Ig* indicates coding sequences and switch recombination sequences of *Ig*. *EcoRI* (E), *BglII* (B), *XbaI* (X). B, Southern blot of DNA samples from maternal (M) and paternal (P) hemizygous carriers hybridized with the RSV probe. DNA samples were digested with *EcoRI* to isolate the *pBR/RSV* region, followed by digestion with either *HpaII*, *HhaI*, *BstUI*, or *MspI*. The 2.5 kb *EcoRI* fragment is enlarged below and *MspI*, *HhaI*, and *BstUI* methylation-sensitive restriction sites are shown as vertical lines of different sizes. C, Southern blot of DNA samples from maternal (M) and paternal (P) hemizygous carriers hybridized with the C \square probe. DNAs were digested with *EcoRI* and *BglII* to isolate the *IgA* region immediately adjacent to the *pBR/RSV* region, followed by digestion with either *HpaII*, *HhaI*, *BstUI*, or *MspI*. The 3.3 kb *EcoRI/BglII* transgene fragment is enlarged below and *MspI*, *HhaI*, and *BstUI* methylation-sensitive restriction sites are shown as vertical lines of different sizes. The C \square probe recognizes a 3.3 kb transgene fragment and an 8 kb endogenous band. All DNA samples were obtained from tail biopsies performed at the time of weaning (3 to 4 weeks). Figure legend adapted from Reinhart et al. (2002).

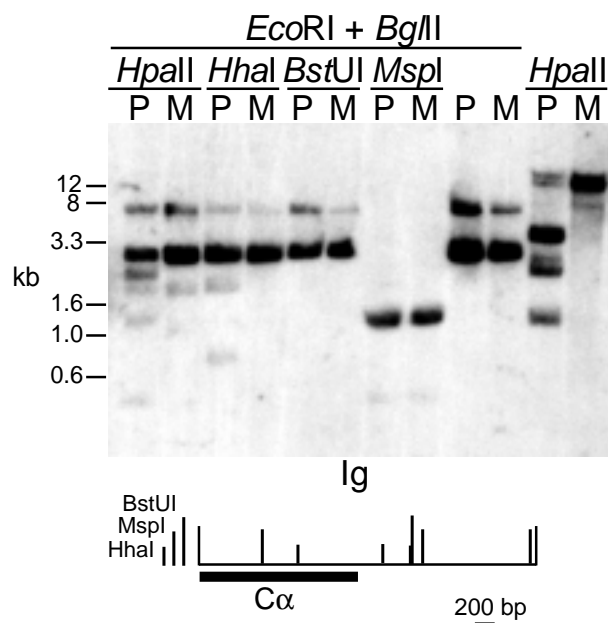
A



B



C



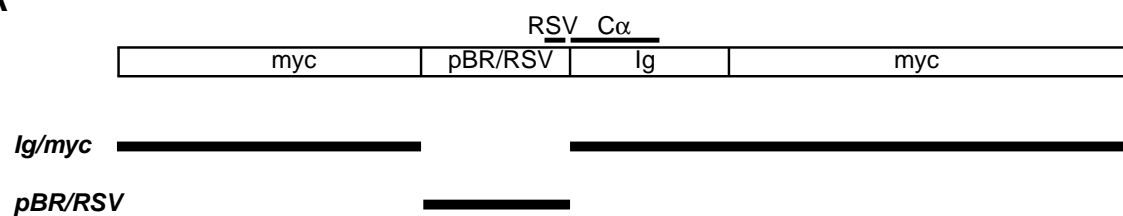
The *Ig* region of the transgene is immediately adjacent to the pBR/RSV region, and includes the C α and S α regions from the *Ig* locus (Figure 12A). The methylation within this region was examined using the same DNA samples and methylation sensitive restriction endonucleases described above (Reinhart et al. 2002). A 3.3 kb region was isolated using the *Eco*RI site adjacent to C α and the *Bgl*III site within S α . DNAs were then digested with the methylation sensitive enzymes *Hpa*II, *Hha*I, or *Bst*UI and hybridized with the C α probe (Figure 12C). After both maternal and paternal inheritance of the *RSVlgmyc* transgene, the predominant band seen on the Southern blot was the uncut, 3.3 kb band. Only minor differences in methylation were observed at *Hpa*II and *Hha*I sites, indicated by the appearance of faint, low molecular weight bands only after paternal inheritance. These data illustrate that the *Ig* region of the transgene is not differentially methylated. It was previously shown that the CpG island located in intron 1 of *c-myc* is not differentially methylated (Howell et al. 1998).

The above data demonstrate that the pBR/RSV region of the *RSVlgmyc* transgene is its differentially methylated domain. Only the pBR/RSV sequences within the transgene are differentially methylated. The sequences surrounding pBR/RSV have equivalent levels of methylation on both parental alleles. The presence of a well-defined DMD within the transgene locus is similar to what is seen at an endogenous imprinted locus.

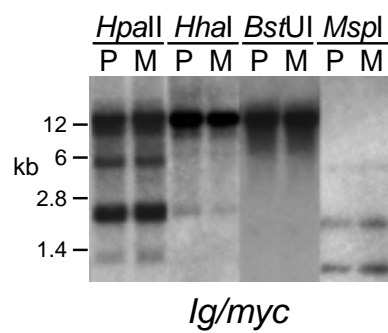
Figure 13. Requirement of the *RSVlgmyc* DMD for transgene imprinting

A, Schematic of transgenes derived from the *RSVlgmyc* sequence. Black lines indicate the probes used in the Southern blots below. *Ig/myc* retains all of the *RSVlgmyc* transgene sequence, except the *pBR/RSV* region between the two *EcoRI* (E) sites. *pBR/RSV* is composed entirely of the sequence delineated by the two *EcoRI* sites. B, Southern blots of DNAs from maternal (M) or paternal (P) hemizygous carriers of the *Ig/myc* transgene digested with either *HpaII*, *HhaI*, *BstUI*, or *MspI* and hybridized with the C \square probe. C, The Southern blot was performed as described above and hybridized with a probe to the RSV region of the transgene. All DNA samples were obtained from tail biopsies performed at the time of weaning (3 to 4 weeks). Figure legend adapted from Reinhart et al. (2002).

A



B



C

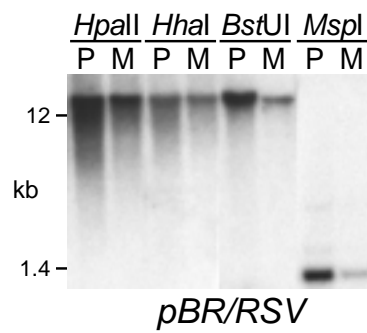


Table 6. Summary of *pBR/RSV* and *Ig/myc* transgenes

Two additional non-imprinted *pBR/RSV* lines were previously analyzed (Chaillet et al. 1995). The number of transgene copies per diploid genome was estimated by comparing the intensity of the transgene hybridization band to the intensity of the hybridization band from the endogenous *Immunoglobulin* heavy chain locus on a Southern blot hybridized with the C α probe. Transgene methylation was determined by comparison of the methylation patterns of maternal and paternal alleles from Southern blots. Results are described as a difference or equivalence between maternal and paternal methylation.

Transgenic Line	Number of transgene copies	Methylation
<i>pBR/RSV A</i>	10	Maternal = Paternal
<i>pBR/RSV C</i>	10	Maternal = Paternal
<i>Ig/myc A</i>	12	Maternal = Paternal
<i>Ig/myc B</i>	10	Maternal = Paternal
<i>Ig/myc C</i>	10	Maternal = Paternal
<i>Ig/myc D</i>	15	Maternal = Paternal

3.3. Deletion of the *RSVlgmyc* DMD abolishes imprinting

At many endogenous imprinted loci deletion of the DMD abolished imprinting (Thorvaldsen et al. 1997; Zwart et al. 2001; Fitzpatrick et al. 2002). Also, for several endogenous imprinted genes the DMD alone, when used as a transgene, was not imprinted (Wutz et al. 1997; Cranston et al. 2001). As described above, transgenes that lacked the pBR322 sequences were maternally methylated and paternally undermethylated. Similarly, transgenes that lacked the RSV LTR sequences were maternally methylated and paternally undermethylated (Chaillet et al. 1995). Thus, neither the RSV LTR sequence alone, nor the pBR322 sequence alone were required for transgene imprinting. However, we have now established that both regions are included within the transgene's DMD. To determine if the entire DMD is required to generate an imprint, the pBR/RSV region of the transgene was removed using flanking *EcoRI* restriction sites. This generated two DNA fragments, *pBR/RSV* and *Ig/myc*, that were used as transgene constructs for the following analysis (Figure 13A). These experiments were performed by Mariam Eljanne.

The *Ig/myc* transgene was tested for its ability to create an imprinted locus. Four independent transgenic lines were generated and the methylation on maternal and paternal *Ig/myc* alleles was compared by Southern blot analysis. DNAs from hemizygous carriers of the *Ig/myc* transgene were digested with *HpaII*, *HhaI*, or *BstUI* and hybridized with the C \square probe (Reinhart et al. 2002) (Figure 13A). Equivalent levels of methylation were present on both parental alleles of the transgene (Figure 13B). The same result was observed in four independent transgenic lines (Table 6). Therefore, the DMD of *RSVlgmyc* is required for imprinting.

The ability of the pBR/RSV region of the transgene to create an imprint was addressed by a previous *RSVlgmyc* deletion construct (Chaillet et al. 1995, and Figure 11, construct *J*). The *pBR/RSV* transgene was not imprinted in two transgenic lines. Maternal and paternal transgene

alleles showed equivalent levels of methylation. These results were confirmed by two additional transgenic lines (Reinhart et al. 2002) (Figure 13C and Table 6). The maternal and paternal transgene alleles of the *pBR/RSV* transgene were compared by Southern blot analysis. A high molecular weight band was observed for both the maternal and paternal transgene alleles. Thus, the *pBR/RSV* transgene was not differentially methylated.

The above data demonstrate that the pBR/RSV sequences comprise the DMD of the *RSVlgmyc* transgene. The pBR/RSV DMD sequences are required for *RSVlgmyc* imprinting. Removal of these sequences leads to a loss of differential methylation. However, the DMD is not sufficient to generate an imprinted locus alone. The pBR/RSV transgene alone is not imprinted. Therefore, the combination of the *Ig/myc* and *pBR/RSV* sequences is necessary to establish an imprint.

3.4. Use of the *RSVlgmyc* transgene as a model to assess DMD sequence function

The DMD of a typical imprinted gene includes a small, well-defined region of the gene. Removal of an imprinted gene's DMD leads to a loss of imprinting. Furthermore, DMDs alone are not able to generate an imprinted transgene locus. These similarities suggest that all DMDs share a common, essential imprinting function. Assuming that all maternally methylated DMDs share a common function, it is reasonable to propose that this function is interchangeable. The *RSVlgmyc* transgene was used as a model system to test this hypothesis.

As described above, *EcoRI* sites delineate the DMD of the *RSVlgmyc* transgene. Removal of the DMD (*Ig/myc* transgene) leads to a loss of imprinting. The assumption that all DMDs share a common function was tested by replacing the *RSVlgmyc* DMD with sequences from other endogenous imprinted gene DMDs. If DMDs do share a common function, exchanging DMD sequences should restore *Ig/myc* imprinting. Hybrid transgenes were constructed by first removing the pBR/RSV region of *RSVlgmyc* with the flanking *EcoRI*

restriction sites. Specific imprinted gene sequences were PCR amplified from genomic DNA with PCR primers designed to introduce flanking *EcoRI* sites. Resulting *EcoRI* fragments were introduced into the unique *EcoRI* site of the *Ig/myc* transgene. Hybrid transgene constructs were used to generate transgenic mice. The ability of a transgene to generate an imprint was evaluated by the presence of parent-specific methylation. Expression from the transgene was eliminated by removal of the RSV sequence and could not be assessed.

Sequences from the *Igf2r*, *Snrpn*, and *H19* imprinted gene DMDs, as well as the long terminal repeat (LTR) region of an IAP element, were used to replace the DMD of *RSVlgmyc* (Figure 14). All hybrid transgenes were analyzed for parent-specific methylation by Southern blot analysis (Reinhart et al. 2002). Tail DNAs from hemizygous carriers of a maternally or paternally inherited transgene were collected at the time of weaning (3-4 weeks). DNAs were digested with the methylation sensitive restriction endonuclease *HpaII*, and Southern blots were hybridized with the C \square probe. A high level of methylation was detected as a high molecular weight, uncut band on a Southern blot. A low level of methylation appeared as a series of low molecular weight *HpaII* restriction fragments.

3.4.1. *Igf2r* DMD sequences functionally replace the *RSVlgmyc* DMD

The *Igf2r* gene is located within an imprinted gene cluster on chromosome 17 (Figure 6). A maternally methylated DMD is located within the second intron of *Igf2r* (DMD2). The DMD is approximately 3 kb in size, and contains the promoter for a paternally expressed, untranslated RNA (*Air*) (Figure 14). The sequence chosen for this analysis is 667 bp in size and includes tandem repeats located downstream of the *Air* promoter (Figure 15A). The sequence was subcloned into the *Ig/myc* transgene to generate the *Igf2rIgmyc* transgene. By Southern blot analysis all maternally inherited *Igf2rIgmyc* transgene alleles were highly methylated, and all

paternally inherited alleles were undermethylated (Reinhart et al. 2002) (Figure 15B). The same result was observed in three independent transgenic lines (Table 7). This pattern of methylation is similar to what is seen with the *RSVlgmyc* transgene. Therefore, the hybrid *Igf2rIgmyc* transgene is imprinted.

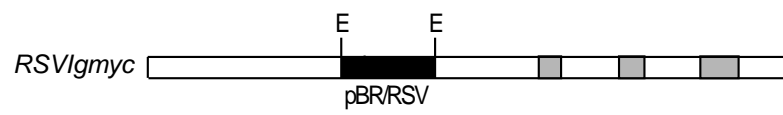
The *Igf2rIgmyc* transgenic lines were generated in an inbred FVB/N genetic background. The imprinted characteristics of the *RSVlgmyc* transgene were consistent when maintained in this background. However, imprinting of the *RSVlgmyc* transgene was affected by differences in mouse strain background. When the *RSVlgmyc* transgene was crossed into the C57BL/6J genetic background paternal alleles acquired a high level of methylation, similar to that seen on the maternal allele (Chaillet et al. 1995). Endogenous imprinted gene methylation and expression are typically not affected by alterations in strain background.

The *Igf2rIgmyc* transgene showed characteristic differential methylation in an inbred FVB/N background. The *Igf2rIgmyc* transgene was also analyzed in an inbred C57BL/6J genetic background. Mice hemizygous for the *Igf2rIgmyc* transgene locus were back-crossed for three generations to inbred C57BL/6J mice (Taconic, Germantown, NY). The methylation of the transgene after maternal and paternal inheritance was compared by Southern blot analysis (data not shown, Reinhart et al. 2002). Unlike the original *RSVlgmyc* transgene, the *Igf2rIgmyc* transgene maintained its imprinting in the C57BL/6J genetic background. Maternal alleles contained a high level of methylation and paternal alleles contained a low level of methylation. These data demonstrate that *Igf2rIgmyc* imprinting is not affected by genetic background.

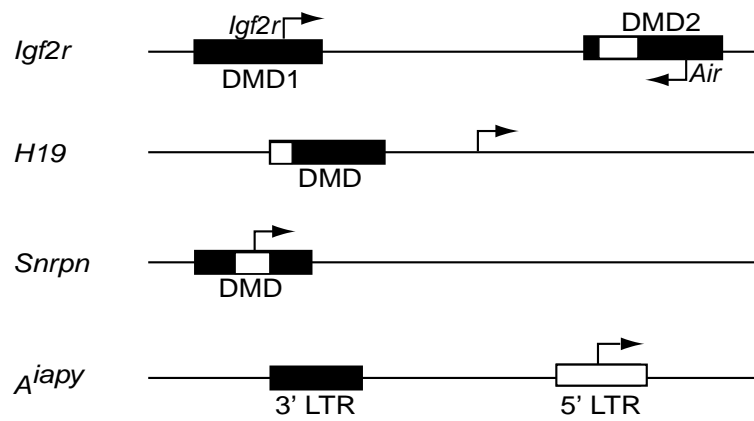
Figure 14. Design of hybrid transgenes

A, Simplified schematic of the *RSVlgmyc* transgene described in Figure 12A. The black rectangle represents the DMD of *RSVlgmyc* (pBR/RSV) flanked by *EcoRI* sites (E). B, Schematics of the *Igf2r*, *H19*, and *Snrpn* imprinted loci, and the *A^{iapy}* IAP element insertion at the *agouti* locus (drawings not to scale). Black boxes represent the DMDs of each locus (or the IAP LTR), and arrows indicate transcription start sites. White portions of each DMD represent the sequences used to create hybrid transgene constructs. C, Schematic of the hybrid transgene constructs created by replacement of the pBR/RSV region of the *RSVlgmyc* transgene with the shaded portions of the DMDs illustrated in panel B. The following sequences were used: *H19*, U19619 nt 1434-1726; *Igf2r*, L06446 nt 741-1408; *Snrpn*, AF130843 nt 3237-3745; IAP reference Michaud et al. 1994 figure 2).

A



B



C

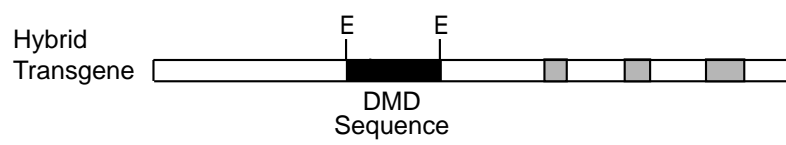
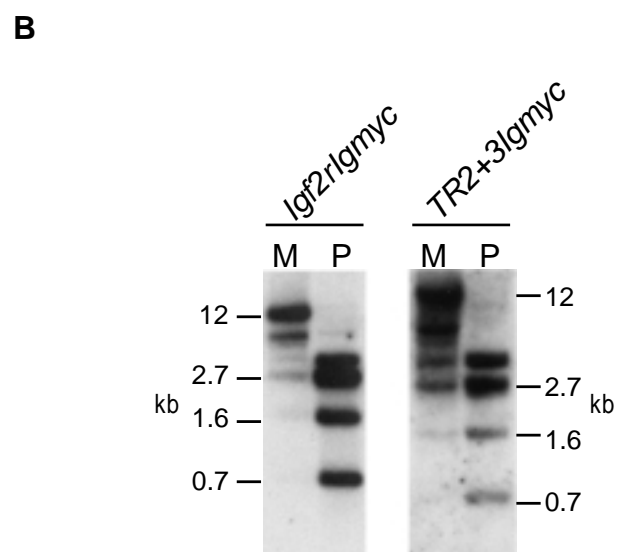
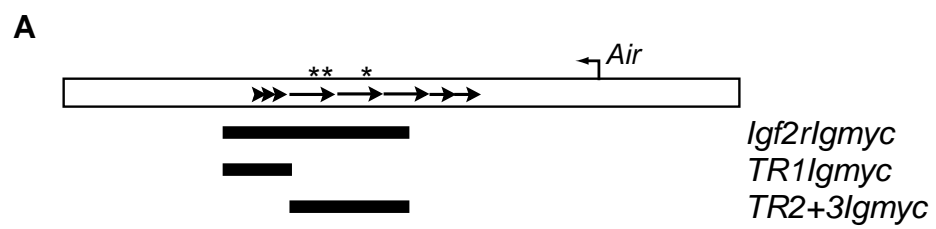


Figure 15. Restoration of transgene imprinting

A, Schematic of the *Igf2r* DMD used to generate the *Igf2rIgmyc*, *TR2+3Igmyc* and *TR1Igmyc* transgenes. Data for the *Igf2rIgmyc* and *TR2+3Igmyc* transgenes are shown. The *Igf2r* DMD sequence is shown by a white rectangle. Transcription start site for the untranslated *Air* RNA is shown as a bent arrow. Arrows within the DMD represent tandem repeats described in Neumann et al. (1995); arrowheads (TR1 repeats), long arrows (TR2+3 repeats), short arrows (TR2+4 repeats). Asterisks mark the location of *HpaII* sites. Thick black lines below the DMD indicate sequences used to generate each transgene. Names of each transgene construct are shown to the right. B, Southern blot analysis of the *Igf2rIgmyc* and *TR2+3Igmyc* transgenic lines. DNAs from maternal (M) or paternal (P) hemizygous transgenic carriers were digested with the methylation sensitive enzyme *HpaII*. DNAs were probed with the C \square probe. All DNAs were obtained from tail biopsies at the time of weaning (3 to 4 weeks). DNA sizes shown in kilobases (kb).



The *Igf2r* DMD contains three sets of tandem repeats classified as TR1, TR2+3, and TR2+4 (Neumann et al 1995). The *Igf2r* sequence within the *Igf2rIgmyc* transgene includes three unit copies of the 30 bp TR1 repeat and two and a half unit copies of the 175 bp TR2+3 repeat (Figure 15A). To determine if either set of tandem repeats alone could restore imprinting to the *Ig/myc* transgene they were independently PCR amplified and introduced into the *Ig/myc* transgene (Reinhart et al. 2002). The *TR2+3Igmyc* transgene contains approximately 2.5 unit copies of the TR2+3 repeat. The *TR2+3Igmyc* transgene was used to generate one transgenic line. Southern blot analysis showed that maternal alleles of the *TR2+3Igmyc* transgene were methylated, and paternal alleles were undermethylated (Figure 15B). The *TR1Igmyc* transgene contains three unit copies of the TR1 repeat. The *TR1Igmyc* transgene was used to generate one transgenic line. Southern blot analysis showed that both parental alleles of this transgene acquired an equivalent level of methylation (data not shown). Therefore the TR2+3 repeats, but not the TR1 repeats, are able to restore imprinting to the *Ig/myc* transgene.

The ability of *Igf2r* DMD sequences to functionally replace the DMD of the *RSV Igmyc* transgene suggests that maternally methylated DMD sequences share a common imprinting function. Furthermore, the imprinting of the hybrid *TR2+3Igmyc* transgene illustrates the capability of the hybrid transgene model system to evaluate the function of specific DMD sequences. The *Igf2rIgmyc* transgene also maintains its imprinting in the C56BL/6 genetic background. These data indicate that hybrid transgenes, composed entirely of mouse genomic sequences, are a good model for an endogenous imprinted gene.

3.4.2. Not all DMD sequences are able to functionally replace the *RSV Igmyc* DMD

The *H19* gene is paternally methylated and maternally expressed. The DMD required for *H19* imprinting is 2 kb in size and is located 2 kb upstream of the *H19* transcription start site (Figure 14). A 292 bp region from the 5' end of the *H19* DMD was introduced into the *Ig/myc* transgene. This sequence is consistently differentially methylated in the gametes, in blastocysts, and in adult tissues (Tremblay et al. 1997; Lucifero et al. 2002). The resulting *H19SIgmyc* transgene was tested for its ability to generate an imprint (Reinhart et al. 2002). The methylation on the maternal and paternal *H19SIgmyc* transgene alleles was compared by Southern blot. Both the maternal and paternal transgene alleles acquired an equivalent pattern of methylation (Figure 16). The same result was obtained from two independent transgenic lines (Table 7). These data demonstrate that this *H19* DMD sequence was not able to restore *Ig/myc* imprinting.

The *Snrpn* gene is located in an imprinted gene cluster on chromosome 7 (Figure 6). The maternally methylated DMD of the *Snrpn* gene is 6 kb in size and includes the promoter, first exon, and first intron of the *Snrpn* gene (Figure 14). A 508 bp region that includes the promoter and first exon of the *Snrpn* gene was introduced into the *Ig/myc* transgene (*SnrpnIgmyc* transgene). This region is consistently differentially methylated in gametes and in adult tissues (Lucifero et al. 2002; Shemer et al. 1997). Southern blot analysis showed that the maternal and paternal alleles of the *SnrpnIgmyc* transgene contained equivalent, low levels of methylation (Figure 16) (Reinhart et al. 2002). The same result was observed in two independent transgenic lines (Table 7). Therefore, the *Snrpn* promoter region is unable to restore imprinting to the *Ig/myc* transgene.

Figure 16. Non-imprinted hybrid transgenes

Southern blot analysis of the *H19SIgmyc* and *SnrpnIgmyc* transgenic lines. DNAs from maternal (M) or paternal (P) hemizygous transgenic carriers were digested with the methylation sensitive enzyme *HpaII* and probed with the C α probe. All DNAs were obtained from tail biopsies at the time of weaning (3 to 4 weeks). DNA sizes shown in kilobases (kb).

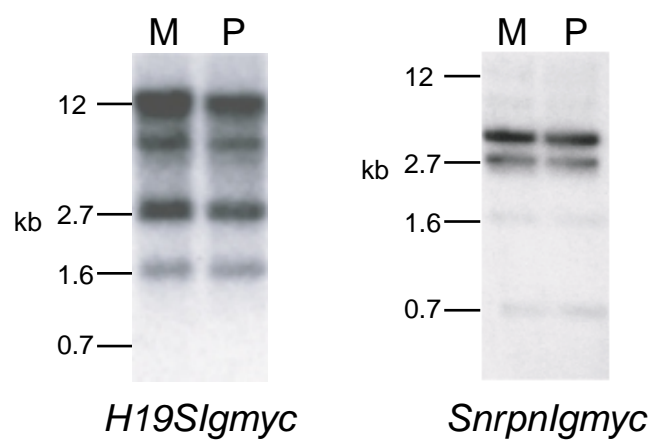


Table 7. Summary of hybrid transgenes

The number of transgene copies per diploid genome was estimated by comparing the intensity of the transgene hybridization band to the intensity of the hybridization band from the endogenous *Immunoglobulin* heavy chain locus on a Southern blot hybridized with the C μ probe. Transgene methylation was determined by comparison of the methylation patterns of maternal and paternal alleles from Southern blots. Results are described as a difference or equivalence between maternal and paternal methylation.

Transgenic Line	Number of transgene copies	Methylation
<i>Igf2rlgmyc A</i>	5	Maternal > Paternal
<i>Igf2rlgmyc B</i>	3	Maternal > Paternal
<i>Igf2rlgmyc C</i>	10	Maternal > Paternal
<i>H19Slgmyc A</i>	10	Maternal = Paternal
<i>H19Slgmyc B</i>	10	Maternal = Paternal
<i>Snrpnlgmyc A</i>	20	Maternal = Paternal
<i>Snrpnlgmyc B</i>	10	Maternal = Paternal
<i>IAPlgmyc A</i>	3	Maternal = Paternal
<i>IAPlgmyc D</i>	15	Maternal = Paternal
<i>IAPlgmyc E</i>	15	Maternal = Paternal
<i>IAPlgmyc F</i>	10	Maternal = Paternal
<i>TR2+3lgmyc</i>	20	Maternal > Paternal
<i>TR1lgmyc</i>	10	Maternal = Paternal

IAP element LTRs are highly methylated on both parental alleles during early development and in somatic tissues. The establishment and maintenance of methylation at IAP LTRs is similar to that of an imprinted gene DMD. A hybrid transgene was generated using the 5' LTR sequence from the *agouti* A^{iapv} allele (Figure 14 and Michaud et al. 1994, figure 2B). The entire 5' LTR sequence was introduced into the *Ig/myc* transgene to generate the *IAPIgmyc* transgene (Reinhart et al. 2002). The *IAPIgmyc* transgene was not imprinted; both parental alleles acquired an identical pattern of methylation, as demonstrated by Southern blot analysis (data not shown). The same result was obtained for four transgenic lines tested (Table 7). Therefore, the IAP element LTR was also not able to restore differential methylation to the *Ig/myc* transgene.

A 292 bp region of the *H19* DMD was not able to establish differential methylation in the *H19SIgmyc* transgene. This could be due to an inability of the small sequence alone to create an imprint. Alternatively, the negative result could indicate that the maternal and paternal germ lines employ different mechanisms to target imprinted genes for methylation. The *H19* gene and the *RSVIgmyc* transgene are normally methylated in opposite germ lines.

Similar to *RSVIgmyc* and *Igf2r*, the *Snrpn* gene is maternally methylated. However, the *Snrpn* promoter sequences in the *SnrpnIgmyc* transgene were not able to establish a maternal imprint. This could have resulted because the *Snrpn* DMD is 6 kb in size and the 508 bp region analyzed was not sufficient to generate an imprint. A larger sequence, or a different sequence may be required. The ability of the *Igf2r* TR2+3 repeat sequences, and not the TR1 repeat sequences, to function as a DMD suggest that not all DMD sequences are equivalent. Therefore, we may be able to identify sequences within the *Snrpn* DMD that are required to create the differential methylation mark.

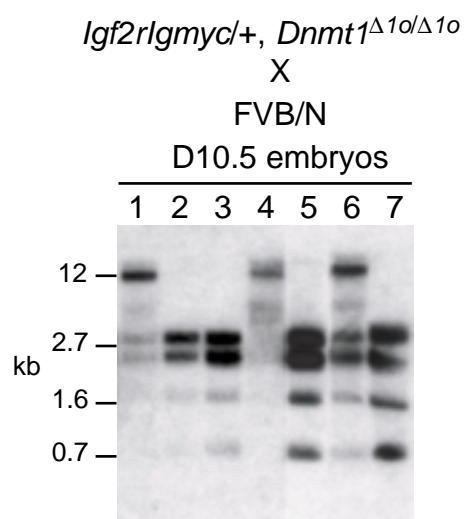
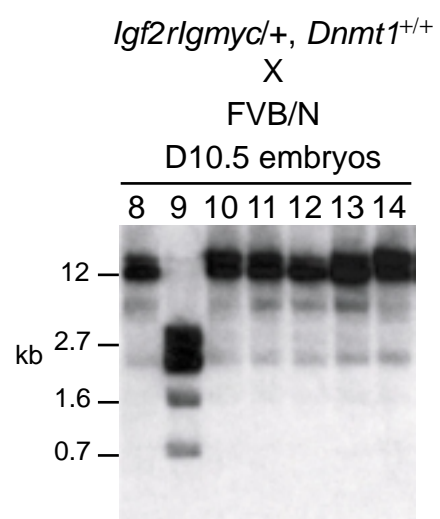
3.5. The *Igf2rIgmyc* hybrid transgene as a model to study genomic imprinting

The *Igf2rIgmyc* transgene closely models many features of endogenous imprinted genes. It is strictly maternally methylated at all sites of integration, and it contains DMD sequences that are required for maternal imprinting. Maternal-specific methylation is seen in both the FVB/N genetic background and the C57BL/6J genetic background. The *Igf2rIgmyc* transgene is a hemizygous locus, present in approximately 5 to 10 copies per haploid mouse genome. These characteristics allow us to easily monitor its parent-specific methylation. These characteristics should also make *Igf2rIgmyc* an ideal model to study the effect of mutations on genomic imprinting. We tested this prediction by studying the effect of the *Dnmt1*^{Δ10} mutation on the methylation of the transgene.

The Dnmt1o methyltransferase is synthesized in the oocyte and is specifically localized in the nucleus of 8-cell stage embryos. The *Dnmt1*^{Δ10} mutation causes loss of Dnmt1o expression from the mutant allele (Howell et al. 2001). Oocytes and preimplantation embryos from homozygous *Dnmt1*^{Δ10/Δ10} female mice contain no Dnmt1o protein. Embryos derived from Dnmt1o-deficient oocytes typically die during the last third of gestation with variable phenotypes and the occasional surviving mouse. Heterozygous embryos generated from Dnmt1o-deficient oocytes show a specific loss of methylation at imprinted loci. Half of the normally methylated alleles of imprinted genes completely lose methylation. These data suggest that Dnmt1o is essential for the maintenance of imprinted gene methylation during the 8-cell stage of preimplantation development.

Figure 17. Loss of Dnmt1o activity affects transgene imprinting

Dnmt1^{fl^o} females carrying the *Igf2rIgmyc* transgene in an FVB/N background (A) or wild-type females carrying the *Igf2rIgmyc* transgene in an FVB/N background (B) were mated to FVB/N males. DNA was extracted from entire transgenic D10.5 embryos, digested with *HpaII*, and Southern blots were performed using the *C* probe to the IgA region of the transgene.

A**B**

Although methylation of the *Igf2r* gene was not examined in heterozygous embryos, we would predict that it would behave as all other imprinted genes tested. To test the effect of loss of Dnmt1o activity on *Igf2rIgmec* transgene methylation, hemizygous carriers of the transgene in the FVB/N genetic background were crossed to mice carrying the *Dnmt1*^{Δ^{lo}} mutation in the FVB/N genetic background. *Dnmt1*^{Δ^{lo}/Δ^{lo}} female mice hemizygous for the transgene locus were mated to *Dnmt1*^{+/+} male mice and embryos were collected at E10.5. Control *Dnmt1*^{+/+} female mice hemizygous for the transgene locus were mated to *Dnmt1*^{+/+} male mice and embryos were collected at the same stage.

Methylation of the transgene was analyzed by Southern blot analysis (Reinhart et al. 2002). Transgenic embryos derived from Dnmt1o-deficient oocytes were compared to wild type transgenic embryos. DNAs were extracted from entire embryos and digested with the methylation sensitive enzyme *HpaII*. Southern blots were hybridized the C⁺ probe. Eight wild type transgenic embryos were analyzed. The majority of embryos showed a pattern of methylation identical to that seen after maternal inheritance of the transgene in adult tail DNA (Figure 17B). One exceptional embryo showed a loss of methylation, yielding a methylation pattern similar to that seen after paternal inheritance of the transgene. Eight transgenic embryos from Dnmt1o-deficient oocytes were analyzed (Figure 17A). Four out of seven embryos showed extensive methylation loss (2, 3, 5, 7). The high molecular weight band indicative of a high level of maternal methylation was completely lost. The pattern of methylation seen was identical to that seen after paternal inheritance of the transgene. The remaining three embryos showed a partial loss of methylation (1, 4, 6). This appeared as a combination of the maternal and paternal

patterns of methylation. Thus, loss of Dnmt1o activity had a profound effect on transgene methylation.

3.6. Analysis of DMD sequences from the paternally methylated *H19* gene

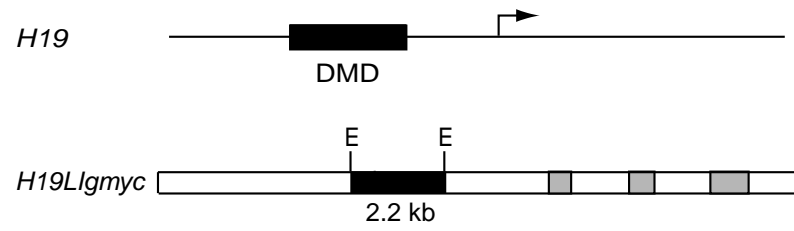
The *H19SIgmyc* transgene was not imprinted. This suggests that the 292 bp sequence analyzed was not sufficient to establish an imprint, or that the *H19* DMD cannot restore imprinting to the transgene. To distinguish between these possibilities, a transgene was constructed that included 2.2 kb of sequence from the *H19* locus. The *H19LIgmyc* transgene contains the entire *H19* DMD in place of the *RSVIgmyc* DMD (Figure 18A). Three *H19LIgmyc* transgenic lines were generated and the maternal and paternal methylation patterns were compared via Southern blot analysis (Figure 18B). The maternal and paternal *H19LIgmyc* alleles contained equivalent levels of methylation. Both parental alleles contained consistent, high levels of methylation in each transgenic line. Therefore, the *H19* DMD cannot replace the DMD of the *RSVIgmyc* transgene.

If sequences from the *H19* DMD share a common function with those of the *Igf2r* DMD they should be included within the sequences analyzed in the *H19LIgmyc* transgene. Not surprisingly, the presence of normally paternally methylated DMD sequences in the context of the *Ig/myc* transgene was not able to restore maternal-specific methylation to the transgene. These data suggest that the imprinting process is markedly different between the two germ lines.

Figure 18. Paternally methylated DMD sequences do not restore transgene imprinting

A, Top, Schematic of the *H19* locus. The 2 kb paternally methylated DMD (black box) of the *H19* locus is located 2 kb upstream of the transcription start site (bent arrow). Bottom, The hybrid *H19Igmyc* transgene was generated by replacing the pBR/RSV region of the *RSVIgmyc* transgene (described in 13A) with 2.2 kb of DNA sequence from the *H19* gene, including the entire DMD. B, Southern blot of DNAs from hemizygous carriers of the *H19Igmyc* transgene. DNAs were digested with *HpaII* and hybridized with the C \square probe. (M) Maternal inheritance of the transgene; (P) Paternal inheritance of the transgene. DNA sizes shown in kilobases (kb).

A



B

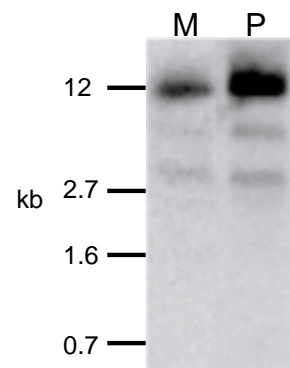
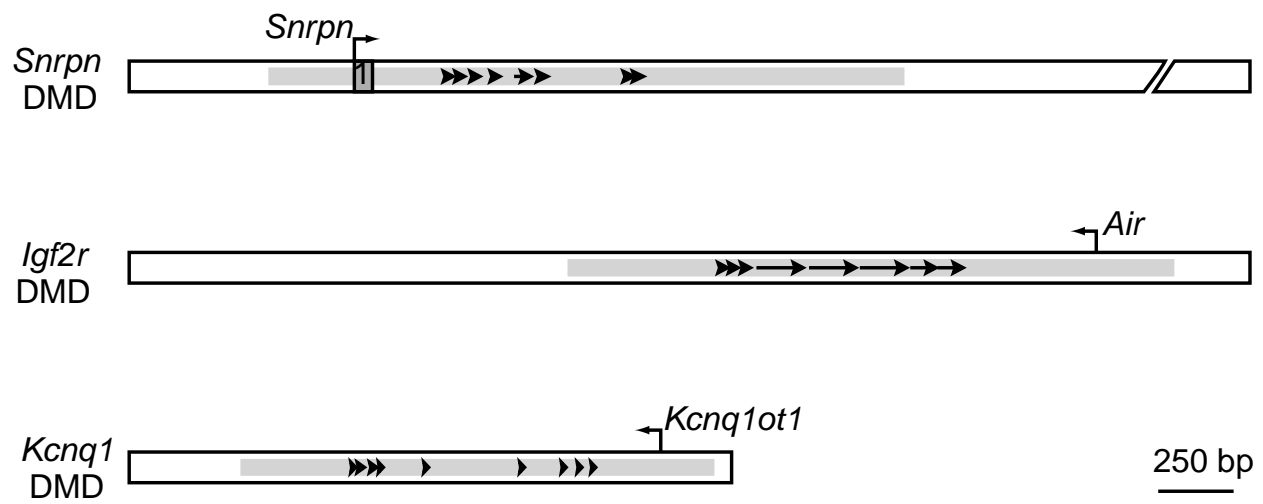


Figure 19. Maternally methylated DMD sequence comparison

Comparison of the DMDs located within the *Snprn*, *Igf2r*, and *Kcnq1* genes. The white rectangles represent the entire DMD sequence (the entire *Snprn* DMD is 6 kb in size). The bent arrows indicate transcription start sites. The *Snprn* DMD contains the transcription start site for the *Snprn* gene and the shaded box represents the first exon of the *Snprn* gene. The *Igf2r* DMD contains the transcription start site for the untranslated *Air* RNA, and the *Kcnq1* DMD contains the transcription start site for the *Kcnq1ot1* untranslated RNA. Gray areas depict CpG islands, and arrows or arrowheads depict the location of tandem repeat sequences (the tandem repeats do not share sequence similarity among the three DMDs).



3.7. Design of hybrid transgenes containing maternally methylated DMD sequences

A shared function of DMD sequences would suggest that they share sequence similarity. However, comparison of the DMDs from various imprinted genes has not revealed any conserved sequences. Moreover, while the methylation differences of the DMD are conserved in the mouse and human there is little sequence conservation. This would suggest that a simple sequence element is not involved in targeting differential methylation to the region. However, certain imprinted genes do share several general characteristics. The maternally methylated DMDs located within the *Snrpn*, *Kcnq1*, and *Igf2r* imprinted genes are compared in Figure 19. These common characteristics were compared to design new hybrid transgenes.

3.7.1. Comparison of endogenous gene DMDs

One similarity between all three regions is their involvement in the coordinate regulation of imprinted gene expression in a large gene cluster (Figure 6). Along with this feature, all three DMDs have several sequence characteristics in common. The DMDs of all three genes contain a CpG island (Shemer et al. 1997; Stoger et al. 1993; Smilnich et al. 2001). Unlike the majority of CpGs islands, those in the DMD regions of imprinted genes are hypermethylated on one parental allele and hypomethylated on the opposite allele. Each DMD also contains the promoter for an imprinted locus. The *Snrpn* promoter is located within the *Snrpn* DMD, the promoter for the *Air* untranslated RNA is located within the *Igf2r* DMD, and the promoter for the *Kcnq1ot1* untranslated RNA is located within the *Kcnq1* DMD (Shemer et al. 1997; Stoger et al. 1993; Wutz et al. 1997; Smilnich et al. 2001). Also, the DMDs of all three genes contain tandem repeats (Neumann et al. 1995; Smilnich et al. 2001; Gabriel et al. 1998). The repeats within each DMD are imperfect repeats, and range in size from 20 to 175 base pairs. They cover

approximately 800 base pair of sequence, and are located within intron sequences of the *Snrpn*, *Igf2r*, and *Kcnq1* genes.

We have shown that the addition of promoter sequences from the *Snrpn* gene to *Ig/myc* did not restore its imprinting. The maternal and paternal alleles were equally methylated. In contrast, addition of tandem repeat sequences from the *Igf2r* DMD to *Ig/myc* restored its imprinting, maternal alleles were methylated and paternal alleles were undermethylated. These findings suggest that tandem repeats are required to establish a maternal imprint. Three transgenes were generated to test the hypothesis that tandem repeats are required to establish maternal-specific methylation. The repeat region from the *Snrpn* DMD and the repeat region from the *Kcnq1* DMD were incorporated into the *Ig/myc* transgene. Also, a single unit copy of the *Igf2r* TR2+3 repeat was incorporated into the *Ig/myc* transgene. If repeated sequences are a requirement for maternal-specific methylation, the *Snrpn* and *Kcnq1* sequences should restore transgene imprinting and the single unit copy of the TR2+3 repeat should not.

3.7.2. Tandem repeats restore imprinting to the *Ig/myc* transgene

An *SnrpnRIgmyc* transgene was created that contains 933 bp of the *Snrpn* DMD including the entire tandem repeat region and a small amount of flanking sequence (Figure 20A). Three *SnrpnRIgmyc* transgenic lines were generated, and the allele-specific methylation patterns of the transgene were compared by Southern blot analysis (Figure 20B). Genomic DNAs from hemizygous carriers of the transgene were digested with the methylation sensitive restriction endonuclease *HpaII* and hybridized with the C \square probe (Figure 20C). The *SnrpnRIgmyc* transgene was imprinted in all three lines (Table 8). Every maternal allele examined was highly methylated, whereas every paternal allele examined was undermethylated.

Two *Kcnq1Igmyc* transgene constructs were created that contain 945 bp of the *Kcnq1* DMD, including the entire repeat region and a small amount of surrounding sequence (Figure 20A). The two constructs contained the repeat region in different orientations with respect to the *Ig/myc* sequences (Table 8). One transgenic line was generated with the *Kcnq1* sequence in one orientation (*Kcnq1Igmyc-1*), and two transgenic lines were generated with the *Kcnq1* sequence in the opposite orientation (*Kcnq1Igmyc-2* and *Kcnq1Igmyc-3*). The methylation of the maternal and paternal transgene alleles was compared by Southern blot (Figure 20B). The *Kcnq1Igmyc* transgene was imprinted in each transgenic line tested (Figure 20C) (Table 8). The maternal alleles were methylated and the paternal alleles were undermethylated. However, in the *Kcnq1Igmyc-2* transgenic line approximately half of the maternal alleles were undermethylated. This was not seen for either of the other transgenic lines.

The *TR2+3SIgmyc* transgene was created that contained one unit copy of the 175 base pair TR2+3 repeat (Figure 21A). The central repeat from the three TR2+3 repeats found at the *Igf2r* locus was chosen for this analysis (Figure 21B). Two transgenic lines were generated and methylation of the transgene after maternal and paternal inheritance was measured by Southern blot (Figure 21C). The *TR2+3SIgmyc* transgene was not differentially methylated. In each line tested the maternal and paternal alleles acquired an equivalent level of methylation. In one transgenic line slight differences in transgene methylation are indicated by slight variation in the intensity of bands after Southern blot analysis (line 1, Figure 21C). This could indicate that the TR2+3 single unit copy retains a minimal ability to create a maternal-specific methylation pattern on the transgene.

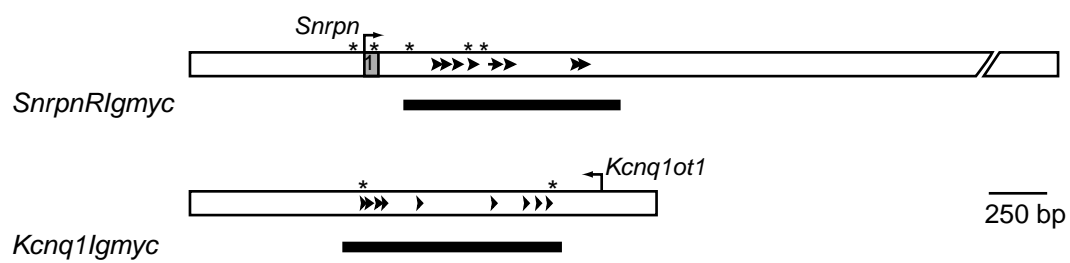
Figure 20. Tandem repeats restore transgene imprinting

A, Schematic of the DMDs from the *Snrpn* and *Kcnq1* genes. Imprinted gene promoters are shown as arrows above the DMD and tandem repeats are shown as arrows within the DMD. Rectangles below each DMD indicate sequences tested by incorporation into the *Ig/myc* transgene. The name of each hybrid transgene is indicated to the left (*SnrpnRIgmyc*, AF130843 nt 3796-4729; *Kcnq1Igmyc*, AF119385 nt 1974-2919). Asterisks denote *HpaII* restriction sites.

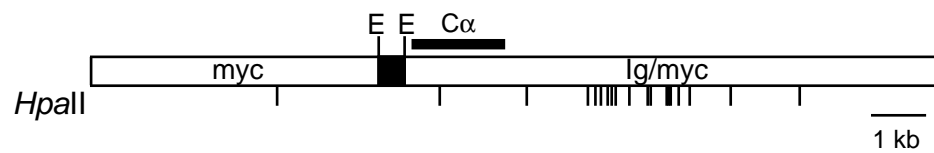
B, Schematic of the *Ig/myc* transgene containing test DMD sequences (black box). The thick black line indicates the location of the C \square probe used for Southern blot analysis. *HpaII* sites are shown as vertical lines below the transgene.

C, Southern blot analysis of genomic tail DNAs digested with the methylation sensitive restriction endonuclease *HpaII* and hybridized with the C \square probe. (M) hemizygous carriers of a maternally inherited transgene. (P) hemizygous carriers of a paternally inherited transgene. DNA sizes are indicated in kilobases (kb).

A



B



C

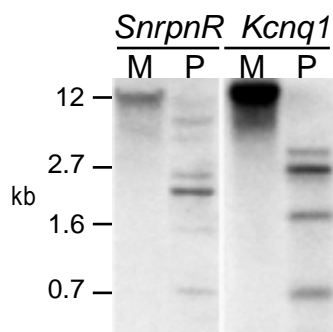
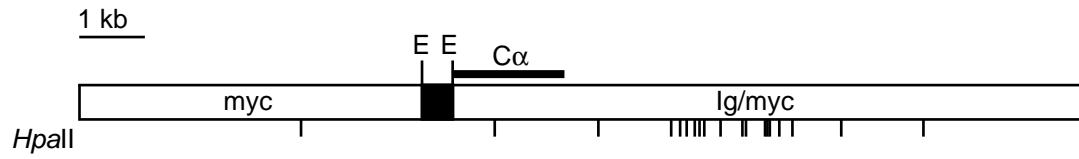


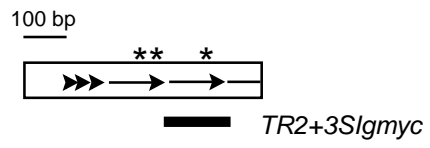
Figure 21. One unit copy of the TR2+3 repeat is not able to restore transgene imprinting

A, Schematic of the *TR2+3SIgmyc* transgene. The TR2+3 DMD sequence is shown as a black box. The thick black line indicates the location of the C \square probe used for Southern blot analysis. *HpaII* sites are shown as vertical lines below the transgene. B, The white rectangle represents the portion of the *Igf2r* DMD that was used to generate the *Igf2rIgmyc* transgene. Arrowheads represent the TR1 repeat and arrows indicate the TR2+3 repeats. Asterisks show the location of *HpaII* sites. The thick black line below the DMD represents the TR2+3 sequence used to generate the *TR2+3SIgmyc* transgene. C, Southern blot analysis of the two *TR2+3SIgmyc* transgenic lines. Genomic tail DNAs were digested with the methylation sensitive restriction endonuclease *HpaII* and hybridized with the C \square probe. (M) hemizygous carriers of a maternally inherited transgene. (P) hemizygous carriers of a paternally inherited transgene. DNA sizes are indicated in kilobases (kb).

A



B



C

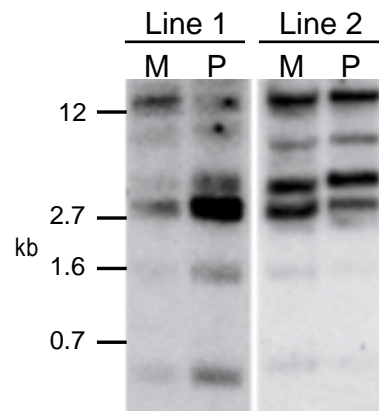
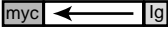
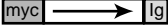
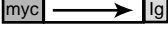
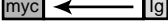
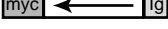
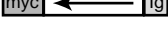
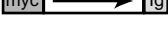



Table 8. Characteristics of hybrid transgenes for analysis of tandem repeats

The number of transgene copies per diploid genome was estimated by comparing the intensity of the transgene hybridization band to the intensity of the hybridization band from the endogenous *Immunoglobulin* heavy chain locus on a Southern blot hybridized with the C μ probe. The orientation of the transgene DMD with respect to the *c-myc* and *Ig* (*Immunoglobulin*) sequences of the transgene is depicted with an arrow. The arrow points in the same direction as the arrows in figure 19. Transgene methylation was determined by comparison of the methylation patterns of maternal and paternal alleles from Southern blots. Results are described as a difference or equivalence between maternal and paternal methylation.

Transgenic line	Number of transgene copies	Orientation	Methylation
<i>Kcnq1lgmyc-1</i>	15		Maternal > Paternal
<i>Kcnq1lgmyc-2</i>	30		Maternal > Paternal
<i>Kcnq1lgmyc-3</i>	15		Maternal > Paternal
<i>SnrpnRlgmyc-1</i>	4		Maternal > Paternal
<i>SnrpnRlgmyc-2</i>	4		Maternal > Paternal
<i>SnrpnRlgmyc-3</i>	30		Maternal > Paternal
<i>TR2+3Slgmyc-1</i>	30		Maternal = Paternal
<i>TR2+3Slgmyc-2</i>	20		Maternal = Paternal

The above data demonstrate that the tandem repeat regions of the *Snrpn* and *Kcnq1* imprinted gene DMDs are able to restore imprinting to the *Ig/myc* transgene. This is similar to the result seen with the *Igf2r* repeat region. These data suggested that the ability of the DMD sequences to restore imprinting to the transgene resides in their tandem repeats. The previously described *Igf2rIgmyc* transgene contained two sets of repeat sequences termed TR1 and TR2+3. The TR2+3 repeats (*TR2+3Igmyc*) were able to restore imprinting to the *Ig/myc* transgene. The *TR2+3Igmyc* transgene contained two complete copies and one partial copy of an imprecise, 175 base pair, repeated sequence. The above results show that one unit copy of the TR2+3 repeat is not able to restore imprinting to the *Ig/myc* transgene.

3.8. Analysis of DMD methylation during development

The data presented in the previous section illustrate the ability to generate hybrid transgenes, composed of endogenous mouse sequences, which are consistently differentially methylated. These transgenes should provide an excellent system to study the methylation of specific DMD sequences during development. Ultimately, we would like to use this system to study the methylation of imprinted genes during preimplantation development. Methylation patterns on the DMDs of imprinted genes are thought to be maintained during preimplantation development, a period when most genomic methylation is lost. However, the maintenance of methylation at each preimplantation stage has not been studied for any imprinted gene. The DMD of the *SnrpnRIgmyc* transgene is consistently imprinted in adult tissues from three transgenic lines. We would therefore expect it to be differentially methylated during preimplantation development.

The above data were obtained by Southern blot analysis with a probe adjacent to the endogenous DMD sequences. A technique more sensitive than the Southern blot needs to be

employed to analyze transgene methylation in preimplantation stage embryos. To examine the methylation of DMD sequences in detail the bisulfite genomic sequencing method was performed. The bisulfite genomic sequencing technique involves treating DNA with sodium bisulfite, which deaminates cytosine to uracil at a defined rate (Figure 22A). Unmethylated cytosines convert more rapidly than methylated cytosines allowing methylated and unmethylated CpGs to be distinguished. PCR primers are designed to amplify the DNA sequence of interest. The PCR products are subcloned and sequenced (Figure 22B). Each sequenced clone contains information about a single allele from the population of alleles within the DNA sample.

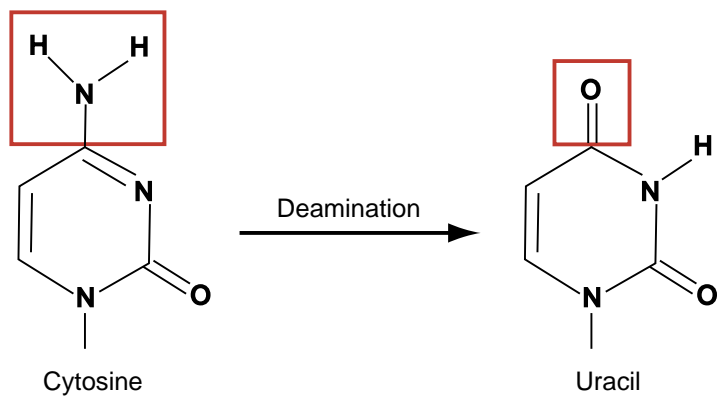
Bisulfite genomic sequencing has advantages over a Southern blot; it can be performed on a smaller amount of DNA, and it adds the ability to examine each CpG within the region of interest. However, this technique also has a disadvantage when examining small amounts of DNA. Small amounts of DNA can lead to bias at the PCR amplification step, possibly leading to the amplification of only one, or a few alleles. This problem can be overcome by using polymorphisms to distinguish the parental alleles of an endogenous gene, and by looking for a large number of sequences containing a single pattern of methylated CpGs (discussed below).

The differential methylation of the *RSVlgmyc* transgene resided in the pBR322/RSV sequences, defining them as its differentially methylated domain. The ability of endogenous DMD sequences to restore imprinting to *Ig/myc* suggests that they are also differentially methylated. The *SnrpnRIgmyc* transgene and the *TR2+3SIgmyc* transgene were chosen for detailed methylation analysis. The imprinted *SnrpnRIgmyc* transgene is expected to have a maternal-specific methylation pattern, and the non-imprinted *TR2+3SIgmyc* transgene is expected to have equivalent levels of methylation on each allele.

Figure 22. The bisulfite genomic sequencing technique

A, Chemical structures of cytosine and uracil. Unmethylated cytosine is deaminated to form uracil by treatment with sodium bisulfite (highlighted by red boxes). B, Diagram of the steps involved in the bisulfite genomic sequencing technique. The DNA is shown as a thick black line. A methylated cytosine in a CpG dinucleotide is highlighted in red and an unmethylated cytosine is highlighted in blue. Sodium bisulfite treatment of DNA leads to the deamination of unmethylated cytosine to form uracil. Methylated cytosine is protected from deamination and forms uracil at a much slower rate. Treated DNA is amplified with PCR primers designed to the DNA region of interest. PCR primers are designed by converting all cytosines (C) to thymines (T) on one strand and all guanines (G) to adenines (A) on the other strand. PCR products are subcloned and sequenced. Unmethylated cytosines in the sequence are now thymines and methylated cytosines remain cytosine.

A



B

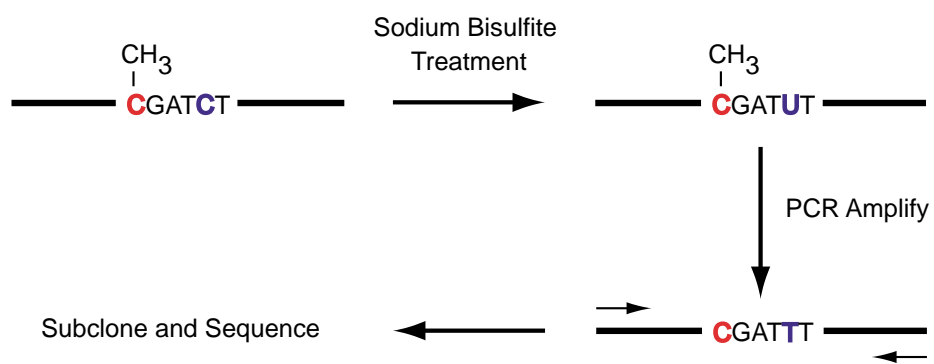
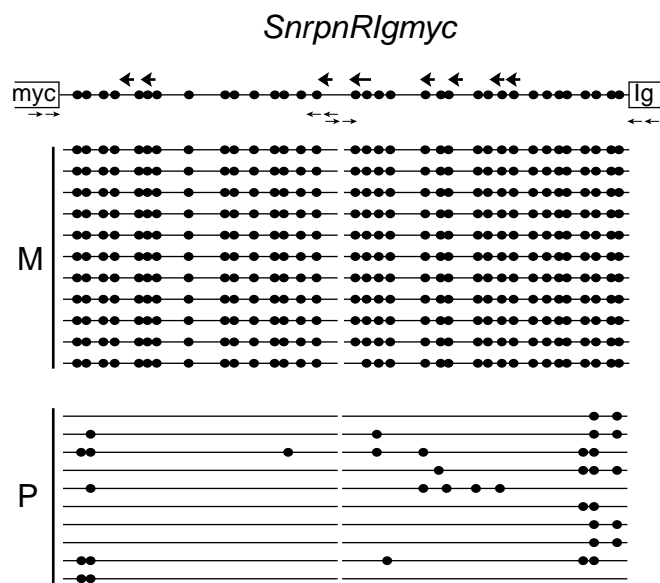


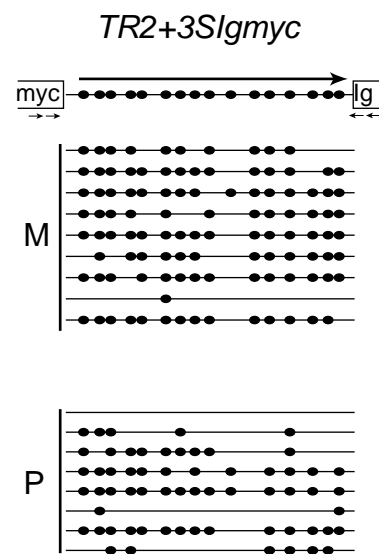
Figure 23. *Snrpn* DMD sequences are differentially methylated in the transgene

A, Schematic of the region of the *SnrpnRIgmyc* transgene analyzed by bisulfite genomic sequencing. The *Snrpn* sequences are shown as a thin line and flanking *Ig/myc* transgene sequences are shown as incomplete white boxes. CpG dinucleotides within the *Snrpn* sequence are shown as filled circles. The repeats within the *Snrpn* sequence are shown as thick arrows (the same orientation as those described in figure 19). Transgene alleles were PCR amplified using two sets of primers (thin arrows). Adult methylation patterns were analyzed using genomic DNA from the spleen of a hemizygous carrier of a maternally (M) or paternally (P) inherited *SnrpnRIgmyc* transgene. Each line indicates a sequenced PCR product. Filled circles represent methylated CpGs. B, Bisulfite genomic sequencing analysis of the *TR2+3SIgmyc* transgene was performed as described for the *SnrpnRIgmyc* transgene with one exception. Only one set of PCR primers was needed to span the entire TR2+3 sequence.

A



B



3.8.1. The *Snrpn* DMD sequences function as *SnrpnRIgmyc*'s DMD

Methylation of the *SnrpnRIgmyc* and *TR2+3Igmyc* transgenes was analyzed in adult tissues by the bisulfite genomic sequencing technique. The *Snrpn* sequence in the *SnrpnRIgmyc* transgene was PCR amplified with two sets of primers that span the entire sequence (Figure 23A). The *TR2+3* sequence was PCR amplified from the *TR2+3SIgmyc* transgene with one set of primers that span the entire sequence (Figure 23B). Adult methylation patterns were analyzed in genomic DNA from the spleen of a hemizygous carrier of a maternally or paternally inherited transgene. Spleen DNA contains the same methylation pattern as tail DNA and was chosen because it is of a higher quality than tail DNA. Each DNA sample was digested with the *HindIII* restriction endonuclease to generate small DNA fragments used for the treatment. Digested DNA was purified and treated with sodium bisulfite. Bisulfite treated samples were PCR amplified, subcloned, and sequenced (Figure 23A). Each maternal *SnrpnRIgmyc* allele was methylated at >95% of CpG dinucleotides, and each paternal *SnrpnRIgmyc* allele was methylated at <22% of CpGs dinucleotides. In contrast, the maternal and paternal *TR2+3SIgmyc* alleles contained similar levels of methylation (Figure 23B). These data demonstrate that the *Snrpn* sequences within the *SnrpnRIgmyc* transgene function as its differentially methylated domain.

3.8.2. DMD methylation is maintained during preimplantation development

The methylation of the *SnrpnRIgmyc* DMD was then analyzed during preimplantation development. Female and male mice hemizygous for the transgene locus were mated to wild type mice, and embryos were collected at different preimplantation stages. Blastocysts stage embryos were flushed directly from the uterine horns. Embryos at the 8-cell stage of development were obtained by collecting embryos at the 4-cell stage and culturing them in CZB

medium to the 8-cell stage. Embryos at the 4-cell stage of development were obtained by collecting 2-cell stage embryos and culturing them in CZB medium to the 4-cell stage. The *in vitro* culture of embryos was done to monitor each embryo and ensure that no embryos went beyond the stage of interest. All culture times were kept as short as possible. Embryos were pooled and analyzed by bisulfite genomic sequencing.

For each sample the region of the *SnrpnRlgmyc* DMD closest to the *c-myc* sequence was analyzed (Figure 23A). This region contains 15 CpG dinucleotides. Ten to twenty blastocyst stage embryos were pooled for analysis (Figure 24). Paternal transgene alleles were always unmethylated with only one allele containing a methylated CpG dinucleotide. Maternal transgene alleles were always methylated. Maternal alleles contained an average of 73% methylated CpG dinucleotides per allele. Only three of 16 tested alleles contained less than 60% methylation. These data clearly illustrate that the *SnrpnRlgmyc* DMD is maternally methylated at the blastocyst stage of preimplantation development.

Methylation at the 8-cell stage of preimplantation development was analyzed in pools of 25 to 30 embryos (Figure 24). Of 11 sequenced paternal alleles eight were completely unmethylated. One allele contained three methylated CpG dinucleotides, and two alleles contained two methylated CpGs. Each maternal allele sequenced was highly methylated. However, out of 10 alleles sequenced only two different patterns of methylated CpGs were detected. There are two possible explanations for the similar patterns of methylated CpGs. It is possible that the same CpG dinucleotides are consistently unmethylated from the pool of 8-cell embryos. However, because small amounts of DNA are obtained from preimplantation embryos, it is possible that the PCR reaction preferentially amplified a few alleles. Regardless of the possible PCR bias, two different patterns of maternal methylation were observed, in comparison

to the unmethylated paternal alleles. These data illustrate that the transgene DMD is differentially methylated at the 8-cell stage.

Methylation at the 4-cell stage of preimplantation development was analyzed in pools of 30 to 50 embryos (Figure 24). All maternal transgene alleles were methylated in the 4-cell stage embryos. Out of 11 sequenced alleles only two different patterns of methylated CpG dinucleotides were observed. Similar to the data obtained for the 8-cell stage sample, this could indicate a PCR bias for only a few maternal alleles. However, the two distinct patterns of methylation still suggest that maternal transgene alleles are methylated at the 4-cell stage. Unfortunately, no data were obtained for the paternal transgene alleles at the 4-cell stage. The PCR was unsuccessful from the pool of 4-cell stage embryos. However, the methylation of maternal alleles at the 4-cell stage suggests that differential methylation of the transgene DMD is maintained at each stage of preimplantation development.

In summary, the *SnrpnRIgmyc* transgene is differentially methylated at all stages of development tested. The *Snrpn* DMD sequences functionally replace the DMD of *RSVIgmyc* and are maternally methylated. The differential methylation of the transgene DMD is maintained at the 8-cell and blastocyst stages. Maternal methylation of the transgene is also seen at the 4-cell stage of preimplantation development. Importantly, these data indicate that methylation on imprinted gene DMDs is maintained during preimplantation development. This has not been clearly shown for any maternally methylated imprinted gene.

3.8.3. Methylation of endogenous imprinted genes in the blastocyst

Only certain imprinted gene DMD sequences are able to function as a DMD in hybrid *Ig/myc* transgenes. For example, the promoter region of the *Snrpn* gene is not differentially methylated in the *SnrpnIgmyc* transgene. However, the *Snrpn* repeat sequences are differentially

methyated in the *SnrpnRlmyc* transgene. This suggests that the ability of DMD sequences to establish differential methylation is not equivalent.

The preimplantation stage of development is an important period for the maintenance of genomic imprints. The above data show that imprinted gene methylation is maintained at this time. However, little is known about the methylation of specific DMD sequences during preimplantation development. It is likely that only certain DMD sequences are differentially methylated during preimplantation. The entire DMD may not be differentially methylated until later in development. To test the function of specific DMD sequences during preimplantation development, the methylation of the promoter and repeat regions of the endogenous *Snrpn* DMD were analyzed at the blastocyst stage of development.

The following experiment was performed using the B6(CAST7) substrain of mice (Mann et al. 2003). The B6(CAST7) mouse genome is primarily C57BL/6J with chromosome 7 from the castaneous strain. The advantage of the B6(CAST7) strain of mice over the castaneous strain of mice is that they are easier to breed. Single nucleotide polymorphisms (SNPs) between the castaneous strain of mice and other inbred strains of mice can be used to distinguish the parental alleles of genes on chromosome 7. SNPs between the castaneous strain and any of the *mus domesticus* strains were estimated to occur once every 200 bps (Lindblad-Toh et al. 2000). A polymorphism in the promoter region of the *Snrpn* gene has already been identified between the castaneous and FVB/N strains of mice. The repeat region of the *Snrpn* locus was PCR amplified from the castaneous or FVB/N genomes and sequenced to identify polymorphisms. One polymorphism was identified and confirmed by PCR and sequencing from the genomic DNA of a heterozygous (FVB/N x castaneous) embryo.

Figure 24. The *SnrpnRlgmyc* DMD is differentially methylated during preimplantation

Bisulfite treatment was performed on embryos embedded in agarose beads as described by Schoenherr et al. (2003). Maternal inheritance of the transgene (M) is compared to paternal inheritance of the transgene (P). For all samples CpG dinucleotides in one half of the *SnrpnRlgmyc* DMD were analyzed. The locations of CpGs are shown as filled circles at the top of each panel. Each sequence is represented by a thin line and methylated CpGs are shown as filled circles.

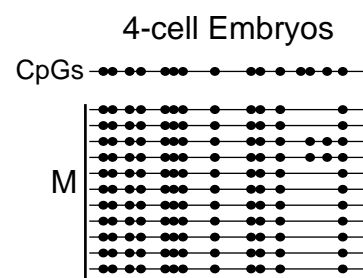
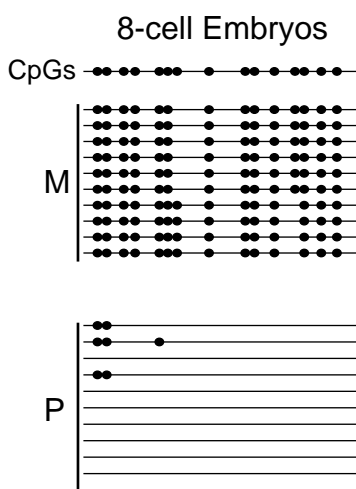
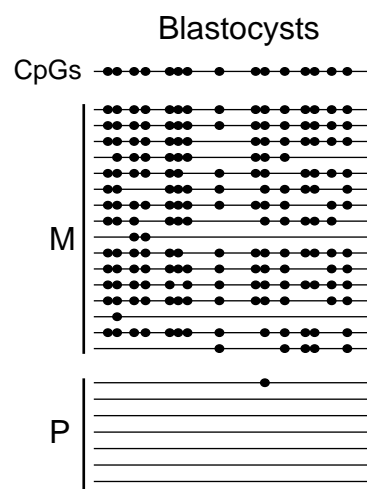
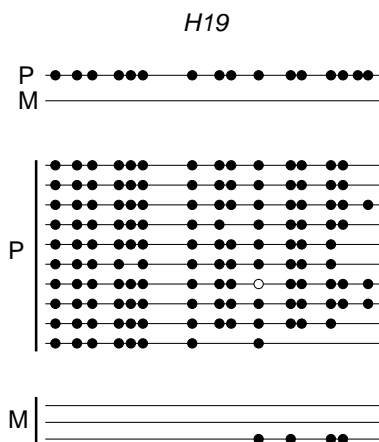
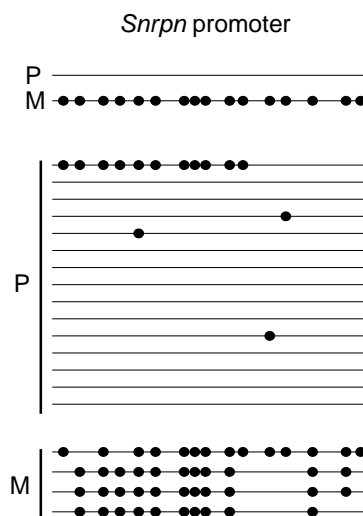
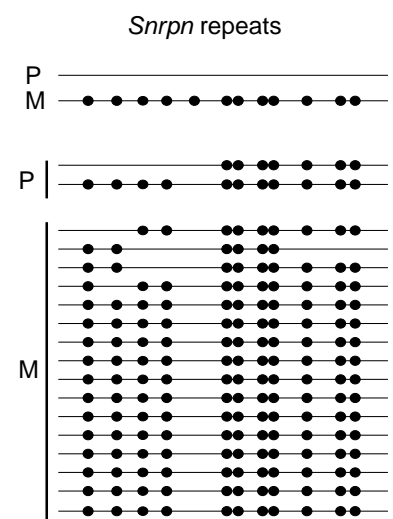


Figure 25. The endogenous *Snprn* gene is differentially methylated in blastocysts

Bisulfite treatment was performed on a pool of 12 blastocysts obtained from a mating of a FVB/N female mouse to a B6(CAST7) male mouse. A, The CpGs at the 5' end of the *H19* DMD were analyzed. The adult methylation patterns on the maternal (M) and paternal (P) alleles of the *H19* gene are shown at the top. The methylation of the same region in the blastocyst sample is shown at the bottom. Methylated CpG dinucleotides are represented by filled circles. White circles indicate untested CpG sites. B, The CpGs in the promoter region of the *Snprn* DMD were analyzed. The adult methylation patterns on the maternal (M) and paternal (P) alleles are shown at the top. The methylation of the same region in the blastocyst sample is shown at the bottom. Methylated CpG dinucleotides are represented by filled circles. C, The CpGs in the repeat region of the *Snprn* DMD were analyzed. The adult methylation patterns on the maternal (M) and paternal (P) alleles are shown at the top. The methylation of the same region in the blastocyst sample is shown at the bottom. Methylated CpG dinucleotides are represented by filled circles.

A**B****C**

Blastocysts were collected from wild type FVB/N females crossed to B6(CAST7) males. Twelve embryos were pooled and subjected to bisulfite treatment. The resulting DNA was PCR amplified with primers to the 5' end of the *H19* DMD, the promoter of the *Snrpn* DMD, and the repeat region of the *Snrpn* DMD. The polymorphisms were used to distinguish the maternal and paternal alleles in each region. The *H19* DMD has previously been shown to be differentially methylated at the blastocyst stage and served as a positive control for the technique (Tremblay et al. 1997). As expected the *H19* DMD was paternally methylated (Figure 25A). Nine out of ten paternal alleles were methylated at > 80% of CpG dinucleotides. Two of three maternal alleles were completely unmethylated, and one allele was methylated at 4 of 15 CpG dinucleotides.

The two regions of the *Snrpn* DMD yielded different results. The promoter region of the *Snrpn* gene contains 16 CpGs that were analyzed for methylation (Figure 25B). Of 15 paternal alleles sequenced, 11 were completely unmethylated. Three alleles were methylated at one CpG and one allele was methylated at 11 CpGs. In contrast, all of the sequenced maternal alleles were methylated. All four alleles were methylated at 10 or more CpGs. The repeat region contains 12 CpG dinucleotides that were analyzed for methylation (Figure 25C). Twelve of 16 maternal alleles sequenced were completely methylated. One allele was unmethylated at one CpG dinucleotide, two alleles were unmethylated at two CpG dinucleotides, and one allele was unmethylated at five CpG dinucleotides. Of the two paternal alleles obtained, one allele was methylated at 11 CpGs, and one allele was methylated at seven CpGs.

These data show that the *Snrpn* promoter region is differentially methylated during preimplantation development. However, differential methylation was not observed for the *Snrpn* repeat region. Paternal and maternal alleles acquired a high level of methylation. These data

suggest that the repeat region of the DMD is not clearly differentially methylated in the blastocyst. However, more samples need to be analyzed to clarify these results.

3.9. Discussion

3.9.1. The mechanisms of maternal and paternal imprinting are distinct

The ability of the *Igf2r* DMD sequences to restore imprinting to *Ig/myc* suggests that imprinted gene DMD sequences share a common function (Figure 15). The results obtained from hybrid transgene containing *H19* DMD sequences suggest that this function is specific to maternally methylated DMD sequences (Figure 16 and Figure 18). Two different *H19* DMD sequences were not able to restore imprinting to the *Ig/myc* transgene. Equivalent levels of high methylation were established on both parental alleles. This finding is not surprising. Imprinted gene sequences are identical between the two germ lines. Methylation must be established specifically on the correct parental allele of an imprinted gene. Separate sets of imprinting sequences and trans-acting factors in the germ lines would ensure that the imprints are correctly established. The inability of other paternally methylated DMD sequences to restore imprinting to the *Ig/myc* transgene would support the idea that imprinting in the maternal and paternal germ lines is performed by different mechanisms.

3.9.2. Analysis of DMD sequence requirements for maternal imprinting

The sequences within a maternally methylated DMD must maintain their parent-specific methylation throughout development. This suggests that sequences within the DMD are required for the establishment of methylation in the oocyte and for the maintenance of methylation during preimplantation. Also, sequences within the DMD must protect the paternal allele from acquiring methylation in the sperm and maintain that protection throughout development. Along with carrying the differential methylation at an imprinted locus the DMD is essential for

imprinted gene expression. The sequences required for these specific functions may be distinct or may overlap. The hybrid transgene analysis described in this chapter examined the requirement of imprinted gene DMD sequences specifically for the establishment and maintenance of a maternal-specific methylation pattern.

The ability of *Igf2r* DMD sequences to functionally replace the DMD of *RSVlgmyc* suggested that the DMDs of other maternally methylated imprinted genes may also possess the same function. To extend this analysis, the DMD sequences of several maternally methylated imprinted gene sequences were compared (Figure 19). The *Snrpn* and *Kcnq1* DMDs, along with the *Igf2r* DMD were chosen for analysis. Each gene is located in an imprinted gene cluster and contains a maternally methylated DMD. Each gene's DMD is required for the imprinted expression of several other imprinted genes in their cluster. This was clearly demonstrated by the results of several deletion experiments (summarized in Figure 6). At each gene cluster, inheritance of a DMD deletion on the normally unmethylated, paternal allele abolished mono-allelic gene expression. Inheritance of the same deletion on the normally methylated maternal chromosome had no effect. Gene expression on the deleted paternal chromosome resembled that of the methylated, maternal chromosome. These data suggest that maternal methylation of the DMD is equivalent to its deletion. A likely role for maternal methylation is to abrogate a gene regulatory activity of the unmethylated DMD.

Comparison of the DMD deletion studies suggests that these three DMDs are involved in common mechanisms of gene regulation. A comparison of the DMD sequences showed that they share several features (Figure 19). Each DMD contains the promoter for an imprinted locus, CpG island sequences, and tandem repeats. The similarities between the DMDs suggest that one or more of these sequence elements are required for imprinting. The promoter

sequences of the *Snrpn* DMD were tested and did not restore imprinting to *Ig/myc* (*SnrpnIgmyc*, Figure 16). These data are consistent with experiments that showed that deletion of the *Snrpn* promoter had no effect on the imprinted expression of genes in the *Snrpn* imprinted gene cluster (Bressler et al. 2001). The *Igf2r* DMD sequences that restored imprinting to the *Ig/myc* transgene contained a portion of its tandem repeats (Figure 15A). This result suggested that tandem repeats may be involved in creating a differential methylation mark. In order to test this hypothesis three transgenes were generated; the *SnrpnRIgmyc* transgene that contained the entire repeat region from the *Snrpn* DMD, the *KcnqIIgmyc* transgene that contained the entire repeat region from the *Kcnq1* DMD, and the *TR2+3SIgmyc* transgene that contained one unit copy of the TR2+3 *Igf2r* repeat (Figures 20A and 21A).

The results from these experiments demonstrated that tandem repeats are required to establish differential methylation at an imprinted gene DMD. Three *SnrpnRIgmyc* transgenic lines showed maternal-specific methylation (Figure 20C and Table 8). The maternal specific methylation of the *SnrpnRIgmyc* transgene was shown to reside within the *Snrpn* DMD sequence (Figure 23). Therefore, while the *Snrpn* promoter sequences were unable to imprint the transgene, sequences 50 bp downstream of the promoter were able to imprint the transgene. At the endogenous *Snrpn* locus both the promoter and the repeats are differentially methylated. However, this analysis clearly shows that they do not possess equivalent functions.

Three *KcnqIIgmyc* transgenic lines also showed maternal-specific methylation (Figure 20C and Table 8). Two different *KcnqIIgmyc* transgene constructs were analyzed that contained the *Kcnq1* DMD sequence in opposite orientations (described in Table 8). Both orientations of the *Kcnq1* DMD restored imprinting to the *Ig/myc* transgene. However, one *KcnqIIgmyc* line was not consistently maternally methylated. Approximately half of the maternally inherited

transgene alleles examined showed a low level of methylation. This could suggest that the imprinting of the transgene is affected by the orientation of the DMD sequence. However, consistent imprinting of a second transgenic line, generated with the DMD sequence in the same orientation, suggests that the occurrence of undermethylated maternal alleles was not due to the orientation of the *Kcnq1* DMD sequence. This effect is most likely due to the transgene integration site and is not a factor of the sequences within the transgene.

Supporting a role for tandem repeats in genomic imprinting the *TR2+3SIgmyc* transgene was not imprinted (Figure 21C). In one transgenic line slight differences in band intensities between the maternal and paternal alleles suggested that the parental alleles might contain slight differences in methylation. However, the overall pattern of bands between the maternal and paternal alleles was the same, and the differences observed were not clear enough to consider the transgene imprinted. These data show that a single unit copy of the TR2+3 tandem repeat is not able to restore imprinting to *Ig/myc*. Still, it is possible that we have merely deleted sequences from the TR2+3 DMD that are required for imprinting. Loss of imprinting may not be due to the loss of repetitiveness. The minimal *TR2+3Igmyc* imprinted transgene will be used as a basis to test the requirement for repetitive sequences in transgene imprinting. A series of transgene constructs were designed that each eliminate a different region of the transgene DMD. Experiments to compare the maternal and paternal methylation patterns of each of these transgenes are currently underway.

The requirement for repeats in the DMD sequence is also supported by the presence of tandem repeats in the DMD of the *RSVIgmyc* transgene. The pBR322 sequences alone are able to establish and imprint, and the RSV sequences alone are able to establish an imprint. Both of these sequences contain tandem repeats. All of the sequences that restore imprinting to the

transgene contain tandem repeats. However, not all repeated sequences are able to restore imprinting to the transgene. The TR1 repeats were not able to establish an imprint in the *TRIIgmyc* hybrid transgene. Moreover, the *Ig/myc* transgene contains direct repeats in the switch (S α) region of the *Ig* sequence and is not imprinted. This suggests that only specific repeat sequences are able to imprint the transgene.

Tandem repeats are also found within the DMDs of other imprinted loci. The imprinted gene *Impact* on chromosome 18 has a repetitive region within its maternally methylated DMD (Okamura et al. 2000). The imprinted gene *Peg10* was recently identified within an imprinted gene cluster on chromosome 6. The maternally methylated DMD at the *Peg10* locus also contains tandem repeats (Ono et al. 2003). The identification of tandem repeats at these other gene clusters suggests that the requirement for tandem repeats in genomic imprinting may be widespread.

These results stress the importance of tandem repeats in the creation of a differentially methylated domain. It has been known for a number of years that direct repeats are associated with imprinted loci. However, this is the first demonstration of a functional role for direct repeats in genomic imprinting. Tandem repeats within DNA sequences are associated with many epigenetic processes other than genomic imprinting. Various forms of gene regulation in yeast, plants, fungi, and mammals have been shown to involve repeated DNA sequences of various lengths and compositions. In many of these processes repeats are the target of DNA methylation, histone modifications, or even point mutations. Typically these alterations lead to the repression of gene expression.

Schizosaccharomyces pombe is subject to gene regulation correlated with the presence of short tandem repeats (Hall et al. 2002). The CenH repeat is a 4.3 kb region containing tandem

repeats similar to the dh centromeric repeats. The CenH repeat region is required for maintaining silencing at the mating type loci, and nucleates methylation of histone 3 at lysine 9 (H3-Lys⁹). This process involves histone deacetylases (HDACs), Clr4 (H3-Lys⁹ methyltransferase), and Swi6 (an HP1 homologue). Interestingly, the centromeric repeats are also capable of silencing gene transcription.

Repeated regions in the *Neurospora* and *Arabidopsis* genomes are also silenced. For example, multiple copies of a transgene introduced into the *Neurospora* genome will be targeted and silenced (Irelan et al. 1994). This process is termed repeat induced point mutation (RIP). The DNA is targeted for both methylation (CpG and non-CpG methylation) and point mutation, or point mutation alone. Several regions of less than 300 bps are capable of acting as methylation signals; however, not all fragments induce methylation at the same level (Selker et al. 1993). Also, repeats alone are not all that is required to trigger *de novo* methylation in *Neurospora*; many repeated transgenes and endogenous duplications are unmethylated (Chang and Staben 1994). In *Arabidopsis* a mechanism similar to RIP, referred to as methylation induced premeiotically (MIP), silences the transcription by DNA methylation alone. The importance of repeated sequences to silencing in this system is illustrated by the fact that if multicopy transgenes are reduced to a single copy by intrachromosomal deletion, they are reactivated (Assaad and Signer 1992).

Silencing of repeated transgene sequences is also observed in the petunia (Linn et al. 1990). In this system, a non-homologous maize gene is expressed when present in a single copy, but is silenced when present in multiple copies (at the same or different genomic locations). Silenced transgenes are hypermethylated when compared to single-copy transgenes. Also, a 1.6 kb repetitive sequence element (RPS) isolated from the petunia genome is able to affect the

transcription of neighboring genes (Ten Lohuis et al. 1995). In silenced regions the RPS element is hypermethylated.

Tandem repeats have recently been shown to correlate with paramutation at the maize *b* locus. Paramutation in maize is an interaction between two alleles where one allele can alter transcription at the other allele. Transcriptional changes are associated with changes in DNA methylation and alterations in chromatin structure (Stam et al. 2002). This alteration is both mitotically and meiotically stable. The paramutagenic B' allele can induce a heritable change in transcription from the paramutable B-I allele. The ability of *b1* alleles to undergo paramutation is dependent upon the number of copies of an 853 bp tandem repeat.

The numerous examples of tandem repeats in other epigenetic processes support a role for tandem repeats in genomic imprinting. In each of these examples, duplication of a sequence leads to its methylation and silencing, suggesting that repetitive DNA sequences are targets for methylation. Repetitive sequences within the mouse genome are also targets for methylation. For example, the mouse *Apvt* gene is targeted for methylation. The targeting of methylation to this locus requires two copies of a mouse B1 element (Yates et al. 1999). Two copies of this element are more efficient at attracting *de novo* methylation than a single copy. Tandem repeats at imprinted loci may perform a similar function. Tandem repeats may attract methylation to the DMD and the maternal-specific pattern of methylation may be achieved in combination with other gene sequences.

At the *SnrpnRlgmyc* locus the combination of *Ig/myc* and *Snrpn* repeat sequences is required for imprinting. The *Snrpn* repeat sequences may target methylation to the transgene in the germ line, and the *Ig/myc* sequences may be involved in establishing the maternal-specific pattern. However, merely targeting methylation to the transgene in the germ line is not sufficient

to generate an imprint. The LTR of an IAP element is targeted for methylation in the germ line, and contains a repetitive region. Hybrid transgenes generated by combining IAP LTR sequences with *Ig/myc* had equivalent, high levels of methylation on both the maternal and paternal alleles (Reinhart et al. 2002). These data suggest that DMD sequences do not function as nonspecific targets of methylation, but that certain sequences are required to achieve a maternal-specific imprint.

3.9.3. Efficiency of transgene imprinting

Maternal alleles of the *Igf2rIgmcy* transgene were always highly methylated in tail DNAs collected from mice at the time of weaning. However, the methylation of maternal *Igf2rIgmcy* alleles at E10.5 was not completely efficient (Figure 17B). One transgenic embryo showed an unmethylated, paternal-like pattern of methylation. This would suggest that methylation was either established and later lost, or was never established. In either case, these data indicate that the efficiency of generating a methylation imprint at the transgene locus is not 100%. It is possible that the minimal DMD sequence present in the *Igf2rIgmcy* transgene, does not allow for completely efficient transgene imprinting.

The data from transgenes containing a single unit copy of the TR2+3 repeat and 2.5 unit copies of the TR2+3 repeat also suggest that efficiency of transgene imprinting is affected by the sequence that is included within the DMD. The *TR2+3Igmcy* transgene (2.5 unit copies of TR2+3) is imprinted, but the methylated maternal alleles do not reach the high levels of methylation seen with the *Igf2rIgmcy* transgene (TR1 repeats plus 2.5 unit copies of TR2+3) (Figure 15B). The data obtained with a single TR2+3 unit copy agree with this idea. Slight differences in methylation are seen between the two parental alleles, however the transgene is clearly not imprinted (Figure 21). These data may demonstrate that multiple copies of the

tandem repeat are required to increase the efficiency of the imprinting process. One unit copy induces minor differences in methylation, 2.5 unit copies restores differential methylation, and combining the TR1 and TR2+3 repeats establishes a high level of maternal-specific methylation. Copy number dependence is seen in other processes involving tandem repeats. The process of paramutation at the maize *b* locus is dependent upon the copy number of a tandem repeat. Also the ability of B1 repeats to target methylation to the *Aprt* locus is more efficient when the sequence is present in two copies (described above).

Decreased efficiency of transgene imprinting was also suggested by the loss of methylation seen on the *Igf2rIgmyc* transgene in embryos obtained from Dnmt1o-deficient oocytes (Section 3.7). Surprisingly, loss of maintenance Dnmt1o activity had a severe effect on transgene imprinting. The transgene lost the majority of its methylation in over half of the embryos examined (Figure 17). At endogenous imprinted loci loss of methylation only occurred at half of the methylated alleles of an imprinted locus, and half of the alleles maintained a wild type pattern. This difference may be attributed to the difference in efficiency of maintenance methylation at the transgene locus compared to an endogenous locus in wild type mice. If imprinting is not 100% efficient at the transgene locus in wild type mice, it is possible that the enhanced effect of Dnmt1o loss is due in part to a general decrease in imprinting efficiency.

3.9.4. Imprinted gene methylation during preimplantation development

An important factor in genomic imprinting is the ability to distinguish the parental alleles of a gene during all stages of development. It has been clearly shown that differential methylation at an imprinted gene DMD is established in the gamete. It has also been well established that methylation is present in the embryo and in the adult. The paternally methylated *H19* gene is clearly differentially methylated in blastocysts (Tremblay et al. 1997). Similarly,

the *RSVlgmyc* transgene is clearly differentially methylated in blastocysts (Chaillet et al. 1995). However, little data exist to describe the methylation of other maternally methylated imprinted genes in preimplantation embryos.

The *SnrpnRIgmyc* transgene was used as a model to study the differential methylation of imprinted gene DMD sequences during preimplantation development. (Figure 24, Section 3.8). These data clearly demonstrate that the transgene is differentially methylated in the 8-cell stage, and at the blastocyst stage of preimplantation development. These data also show that maternal methylation on the DMD is present at the 4-cell stage. These data support a methylation dependent step for imprinting at all stages of early development. However, transgene methylation needs to be examined at the 2-cell stage and on the paternal allele at the 4-cell stage to draw a clear conclusion.

The inability of all DMD sequences to imprint the transgene suggests that not all DMD sequences possess the same functions. As described above, the *Snprn* promoter sequences are not able to restore imprinting to the *Ig/myc* transgene, and the *Snprn* repeat sequences are. However, both of these sequences are differentially methylated in the adult. To test the ability of both of these sequences to maintain methylation during preimplantation development the methylation in both regions was examined at the blastocyst stage. Interestingly, the promoter sequences of the DMD were clearly differentially methylated at this time. However, the repeat sequences of the transgene were not as clearly differentially methylated. The paternal alleles contained a higher level of methylation than expected. This was not the anticipated result for these experiments. The differential methylation of the *Snprn* repeats in the transgene predicted that they would be differentially methylated at the endogenous locus. However, these

experiments do support the fact that not all transgene sequences have equivalent functions with respect to the maintenance of an imprint during development.

The difference in methylation between the transgene data and the endogenous gene data may rest in the fact that slightly different regions of the repeats were analyzed in each experiment. The 3' region of the *Snprn* repeats from the *SnprnRIgmyc* transgene was analyzed in the blastocyst; however, the 5' region of the repeats was not analyzed. The repeats analyzed at the endogenous locus were those in the 5' region of the repeat. These results indicate that a more detailed analysis of methylation at the *Snprn* endogenous locus is required to identify the functional sequences.

Chapter 4: Characterization of IAP element methylation in the blastocyst

4.1. Introduction

The current data suggest that methylation of the IAP LTR is an important form of transcriptional regulation. Data also suggest that LTR methylation is maintained during preimplantation development. Maintenance of methylation during preimplantation is exceptional due to the fact that the majority of genomic methylation is lost by the preimplantation blastocyst stage. This indicates that maintaining LTR methylation is critical. Yet information also suggests that methylation on IAP LTRs decreases in the preimplantation blastocyst. Southern blot analysis of blastocyst DNA hybridized with a probe to the IAP LTR suggested that a *Hpa*II site in the LTR is unmethylated in a small percentage of IAPs at this time (Walsh et al. 1998). Also, bisulfite genomic sequencing of the IAP element LTR showed that IAPs retain only 62% of their methylation in the blastocyst (Lane et al. 2003). These sequencing data suggested that each IAP losses a portion of its methylation, and that only a few IAPs are completely unmethylated. Apart from these data, little is known about the methylation of IAPs during development.

Loss of methylation in the blastocyst may affect IAPs stochastically, or it may affect specific IAPs. One way to determine if specific IAPs are preferentially demethylated in the blastocyst stage is to directly measure the methylation of individual IAP elements in the blastocyst. Characterizing the unmethylated IAPs is of interest because of their ability to escape the host defense system. Escape from methylation would allow active transcription and possibly transposition, ultimately leading to heritable germ line mutations. Learning the characteristics of

this population of IAPs will further our understanding of how sequences are targeted for methylation, and the factors that affect their escape from methylation.

4.2. Methylation of the general IAP element population

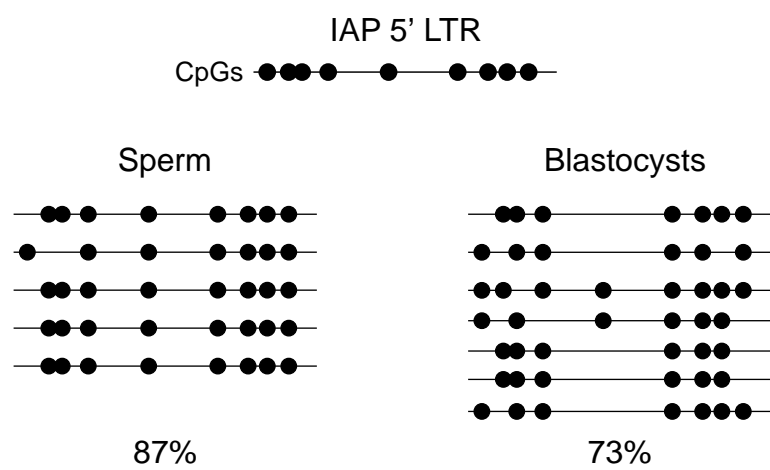
Limited data exist to describe the methylation of IAP elements in the blastocyst, and no information is available concerning the methylation of individual IAPs. In the following experiments the bisulfite genomic sequencing technique was performed to examine the methylation of individual IAP elements in the blastocyst (described in Figure 22). The methylation of the general IAP element population in the blastocyst was evaluated first. This was done to assess the ability of the bisulfite genomic sequencing technique to analyze the methylation of IAPs from a small amount of blastocyst DNA. These data will also confirm the data obtained by others regarding the methylation state of IAP LTRs at the blastocyst stage.

In this experiment PCR primers were designed within conserved regions of the LTR. This PCR will amplify representative LTR sequences from the genome. Blastocysts were collected from wild type FVB/N females crossed to wild type FVB/N males at 3.5 dpc. DNA was collected from blastocysts and treated with sodium bisulfite. Treated DNA was PCR amplified, and PCR products were cloned and sequenced. Each LTR sequence contained nine CpG dinucleotides that were analyzed for the presence or absence of methylation. Data from seven sequenced clones illustrated that the IAP LTR contained a relatively high level of methylation at the blastocyst stage (Figure 26). A total of 73% of CpG dinucleotides were methylated in the blastocyst. No individual LTR sequence was completely unmethylated.

The methylation of IAP LTRs in blastocyst DNA was compared to their methylation in sperm DNA. Sperm DNA is expected to have a higher level of methylation than blastocyst DNA. Sperm DNA was isolated from FVB/N mice and treated with sodium bisulfite. The IAP LTR was PCR amplified with the same primers used for the blastocyst sample. PCR products

Figure 26. IAP element LTRs are methylated in sperm and blastocysts

DNA was collected from a pool of 60 FVB/N blastocysts or from FVB/N sperm by proteinase K treatment. DNA was bisulfite treated and PCR amplified with primers to the conserved region of the IAP 5' LTR (CpGs analyzed within the 5' LTR are shown as filled circles). Each sequenced allele is shown as a thin line and methylated CpGs are shown as filled circles. The percentage of methylated CpGs out of the total number of CpGs sequenced is shown below each sample.



were cloned, and five clones were sequenced. A total of 87% of the CpG dinucleotides analyzed were methylated (Figure 26). Four clones were methylated at eight CpG dinucleotides and one clone was methylated at seven CpG dinucleotides.

These data illustrate that the LTR is predominantly methylated in the mature male gamete. A small amount of LTR methylation is lost by the blastocyst stage of preimplantation development. From the small number of IAP LTRs sequenced in the blastocyst DNA sample none were completely unmethylated. This suggests that the majority of IAP element methylation is maintained during preimplantation development.

4.3. Identification of unmethylated IAPs in blastocyst stage embryos

4.3.1. Experimental design

The above data confirm that the LTRs of the majority of IAP elements are methylated at the blastocyst stage. The goal of these experiments is to identify and characterize unmethylated IAPs at the blastocyst stage. How do we separate the unmethylated IAPs from the hundreds of IAPs present in the mouse genome? Most genomic DNA is unmethylated at the blastocyst stage of development. Therefore, an unmethylated IAP should have an unmethylated LTR in the context of unmethylated flanking genomic sequence. In contrast, the majority of IAPs will contain a methylated LTR in the context of unmethylated flanking sequence. This distinguishing feature was used to screen for unmethylated IAPs in blastocyst DNA.

Blastocyst stage embryos were collected from FVB/N mice. Blastocysts were pooled, and genomic DNA was collected by proteinase K digestion. Embryo DNA was digested with the methylation sensitive restriction endonuclease *HpaII*. There is a conserved *HpaII* site located within the IAP element LTR (Figure 7). *HpaII* digestion will only cut unmethylated IAPs at this *HpaII* site, and at the next neighboring unmethylated *HpaII* site. This digestion will yield a pool of small DNA fragments that contain unmethylated IAP element LTRs. Any methylated LTRs

will remain uncut and produce larger DNA fragment. Digested DNA was ligated to adaptor oligos and PCR amplified. PCR conditions were designed to amplify small DNA fragments containing LTR sequence (one primer designed to the LTR). PCR products were used to generate a lambda library. The library was screened with a probe to the IAP LTR (Michaud et al. 1994).

Thirty clones were selected and sequenced to identify the IAP LTR sequence and flanking genomic sequence. BLAST searches were performed using the DNA sequence flanking the LTR. Clones containing internal IAP genomic sequence were eliminated. A second group of IAP element clones were eliminated because they were located within repetitive regions of the genome. Only four clones contained unique sequences suitable for further analysis. These IAP clones were designated IAP 7, IAP 21, IAP 23, and IAP 29.

Along with analyzing the unmethylated IAPs found in the blastocyst, it is also of interest to examine the methylation of the general IAP element population in the blastocyst. This was previously done on a pool of FVB/N blastocysts using conserved LTR primers (described in section 4.2). However, analyzing IAP LTR methylation at specific genomic insertion sites will offer more valuable information than analyzing LTR methylation using generic LTR sequences. This analysis was performed by randomly choosing IAP elements from the genome. Sequence information for the C57BL/6J genome is easily accessed via the UCSC genome browser (<http://genome.ucsc.edu/>). This database estimates that 90-96% of the genome sequence is known and correct. This web site offers the ability to perform "BLAT" searches against known mouse sequences. BLAT searches use query sequences of greater than 40 nucleotides and locate regions of similarity from the mouse genome sequence. This feature was used to select IAP elements at random from the genome.

Table 9. IAP element characteristics

The IAP elements identified in the blastocyst library screen (IAP 7, IAP 21, IAP 23, and IAP 29), and the IAP elements identified by BLAT search are listed. The chromosomal locations of the IAPs and the mouse strains in which they are found (either FVB/N (F) or C57BL/6J (B)) are provided. The approximate size of each IAP is also described. All information was obtained using the UCSC genome browser (<http://genome.ucsc.edu/>).

IAP Element	Chromosome	Strain FVB/N (F); C57BL/6J (B)	Size
7	10	B/F	5.25 kb
21	11	B/F	7.1 kb
23	13	F	Single LTR
29	13	B/F	Single LTR
30	5	B/F	6.8 kb
31	12	B/F	7 kb
32	18	B	7.2 kb
33	3	B	7 kb
35	12	B	Single LTR
36	1	B/F	7 kb
37	17	B/F	6.2 kb
39	6	B	4.8 kb
40	15	BF	4.4 kb

A BLAT search was conducted using the conserved U3 region of an IAP element LTR (Sequence update February 2003). Sequence from the 5' LTR of the IAP insertion at the *agouti* locus (*A^{iapy}*) was used for the search. IAPs were chosen from different mouse autosomes, and at different locations within or near known or predicted genes. These IAPs were designated IAP 30 through IAP 40. The UCSC BLAT search was also used to identify the genomic location of the IAPs identified in the library screen. The location of each IAP is summarized in Table 9. The methylation at the 5' LTR of each IAP element was analyzed in blastocyst and adult DNA.

4.3.2. Analysis of IAP element methylation

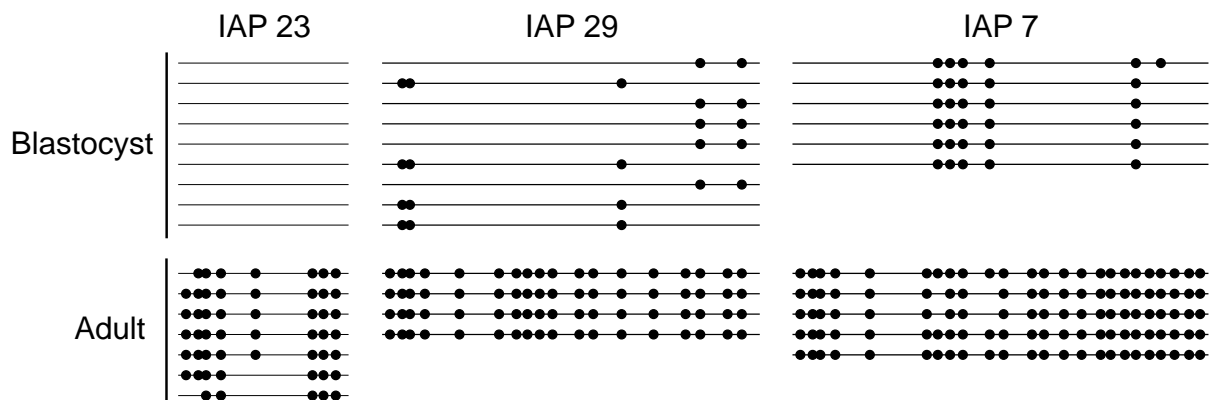
All methylation analyses were performed by the bisulfite genomic sequencing method. PCR primers were designed within the flanking genomic sequence of each LTR. Because the majority of IAP sequences were selected from the C57BL/6J genome, the following analysis was performed using the C57BL/6J strain, unless otherwise stated.

IAP 23 is a single LTR that is only present in the FVB/N strain of mice. Bisulfite genomic sequencing of the IAP 23 LTR was performed on a pool of 40 FVB/N blastocysts. Every IAP 23 LTR sequenced from blastocyst DNA was completely unmethylated (Figure 27A). Each sequenced LTR contained eight CpGs. No methylated CpGs were observed in nine LTRs sequenced. The methylation of IAP 23 was then analyzed in sperm and adult DNAs. Unlike in the blastocyst DNA, IAP 23 was fully methylated in sperm DNA (data not shown). In adult DNA the IAP 23 LTR was also highly methylated (Figure 27A). From 12 sequenced clones only five CpGs were unmethylated. Unmethylated CpGs were located at positions scattered throughout the LTR. No individual LTR contained more than three unmethylated CpGs. Therefore, IAP 23 is specifically unmethylated in blastocysts.

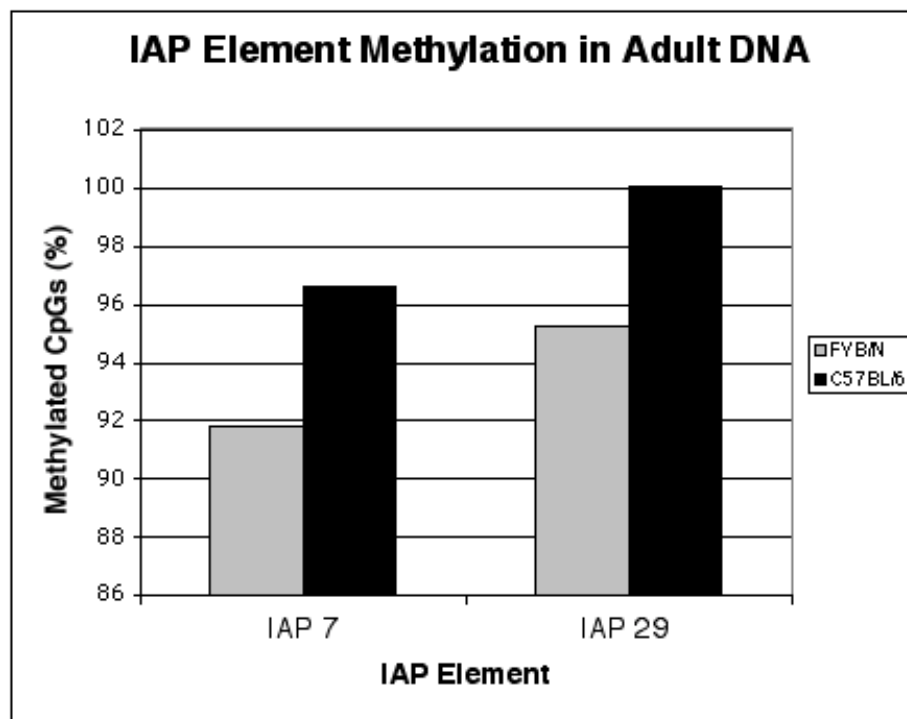
Figure 27. Identification of unmethylated IAP elements from blastocysts

A, Bisulfite genomic sequencing of individual IAP LTRs was performed on pools of FVB/N (IAP 23) or C57BL/6J (IAP 7 and IAP 29) blastocysts. Bisulfite genomic sequencing of adult samples was performed using FVB/N or C57BL/6J tail DNAs. Thin lines represent individual sequenced alleles and filled circles represent methylated CpGs. B, Bisulfite genomic sequencing was performed on FVB/N and C57BL/6J DNAs for IAP 7 and IAP 29. The data obtained are summarized in the chart. The percentage of CpG dinucleotides methylated in each sample is shown on the y-axis. Data for the FVB/N strain are represented by the gray bar and data for the C57BL/6J strain are represented by the black bar (x-axis). The percentage of methylated CpGs reflects the number of methylated CpGs out of the total number of CpG dinucleotides sequenced.

A



B



Similar to IAP 23, IAP 29 is a single LTR insertion. Methylation of IAP 29 was analyzed from a pool of 10 C57BL/6J blastocysts and from adult genomic DNA. The IAP 29 LTR contains 18 CpGs dinucleotides that were analyzed for methylation by bisulfite genomic sequencing. All IAP 29 LTRs examined were unmethylated (Figure 27A). From nine sequenced alleles only 22 CpGs were methylated (14.3% methylation). No more than 3 CpGs were methylated on any one LTR. In contrast, IAP 29 was completely methylated in C57BL/6J adult DNA (Figure 27A). In four sequenced clones no CpG dinucleotides were unmethylated. Therefore, IAP 29 is also specifically undermethylated in blastocysts.

IAP 7 and IAP 21 are both full length IAP elements. The methylation of the IAP 7 5' LTR was analyzed in a pool of 10 C57BL/6J blastocysts and in adult tissue. At the blastocyst stage only 21.5% of the total CpG dinucleotides analyzed were methylated (Figure 27A). Thus, the IAP 7 5' LTR is relatively unmethylated in blastocysts. In contrast, 97% of the total CpG dinucleotides sequenced were methylated in adult DNA. Methylation analysis of the IAP 21 5' LTR was performed using the same pool of 40 blastocysts that was used to analyze methylation of the IAP 23 LTR. Unlike the other three IAPs identified in the library screen IAP 21 was completely methylated in FVB/N blastocysts (data not shown). These data demonstrated that the blastocyst library screen selected for unmethylated IAPs.

In order to determine if there was any strain variability in the methylation of these unique IAPs, the IAP 7 and IAP 29 alleles were analyzed by bisulfite genomic sequencing in C57BL/6J and FVB/N adult DNAs. Both LTRs were consistently less methylated in the FVB/N strain of mice than in the C57BL/6J strain of mice (Figure 27B). The IAP 7 LTR was 92% methylated in the FVB/N strain and 97% methylated in the C57BL/6J strain. Likewise, the IAP 29 LTR was

95% methylated in the FVB/N strain and 100% methylated in the C57BL/6J strain. In both strains, each LTR sequence was methylated. The differences observed between any two LTRs occurred at CpG dinucleotides scattered throughout the LTR. These data indicate that there is little variability in the methylation of IAP LTRs in adult tissues.

The methylation of randomly chosen IAPs was compared to the methylation of those selected from the library screen. One single LTR insertion and seven IAPs were chosen from the C57BL/6J genomic sequence. The methylation of each IAP was analyzed in blastocyst DNA and in adult DNA. Blastocyst methylation analyses were done using two separate pools of 10 C57BL/6J blastocysts. Adult methylation analyses were done using C57BL/6J DNA and FVB/N DNA. Specific primers were designed to amplify the 5' LTR from each IAP.

The selection of two undermethylated, single LTR insertions from the library screen suggested that single LTR insertions are preferentially demethylated in the blastocyst. To test this notion, the methylation of the IAP 35 single LTR insertion was analyzed in blastocyst DNA and in adult DNA. Interestingly, the IAP 35 insertion was undermethylated in both the blastocyst and adult DNAs (Figure 28). Only 13% of the CpG dinucleotides sequenced were methylated on the IAP 35 LTR in blastocyst DNA. The pattern on methylated CpGs was different on each LTR, and no LTR was completely methylated. Each IAP 35 LTR sequenced contained 18 CpGs. In a collection of 10 sequenced LTRs no more than three methylated CpGs were found per LTR. In adult DNA the IAP 35 LTR was also unmethylated. Only 7% of the CpG dinucleotides sequenced were methylated on the IAP 35 LTR. Again, no sequences were completely methylated. Thus, the IAP 35 LTR was undermethylated in both the blastocyst and in the adult.

Figure 28. Single IAP LTRs are unmethylated at the blastocyst stage

IAP 35 is a single IAP LTR chosen from the C57BL/6J genome. Bisulfite genomic sequencing was performed on the IAP 35 LTR from a pool of 10 C57BL/6J blastocysts and from C57BL/6J tail DNA. The filled circles on the top line indicate the positions of CpGs in the IAP 35 LTR. Thin lines represent sequenced clones and filled circles represent methylated CpGs.

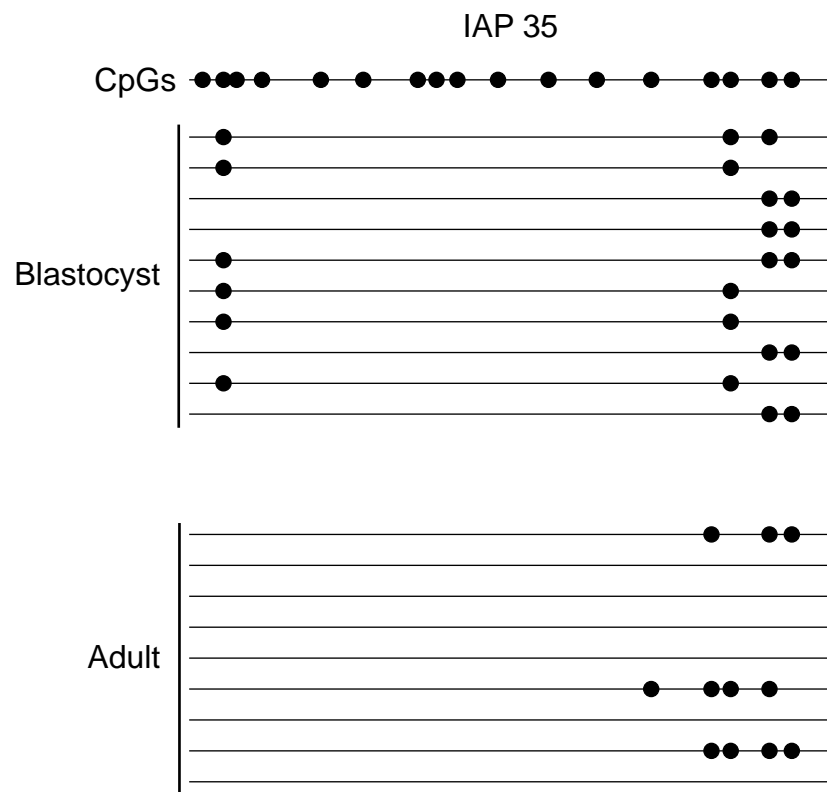
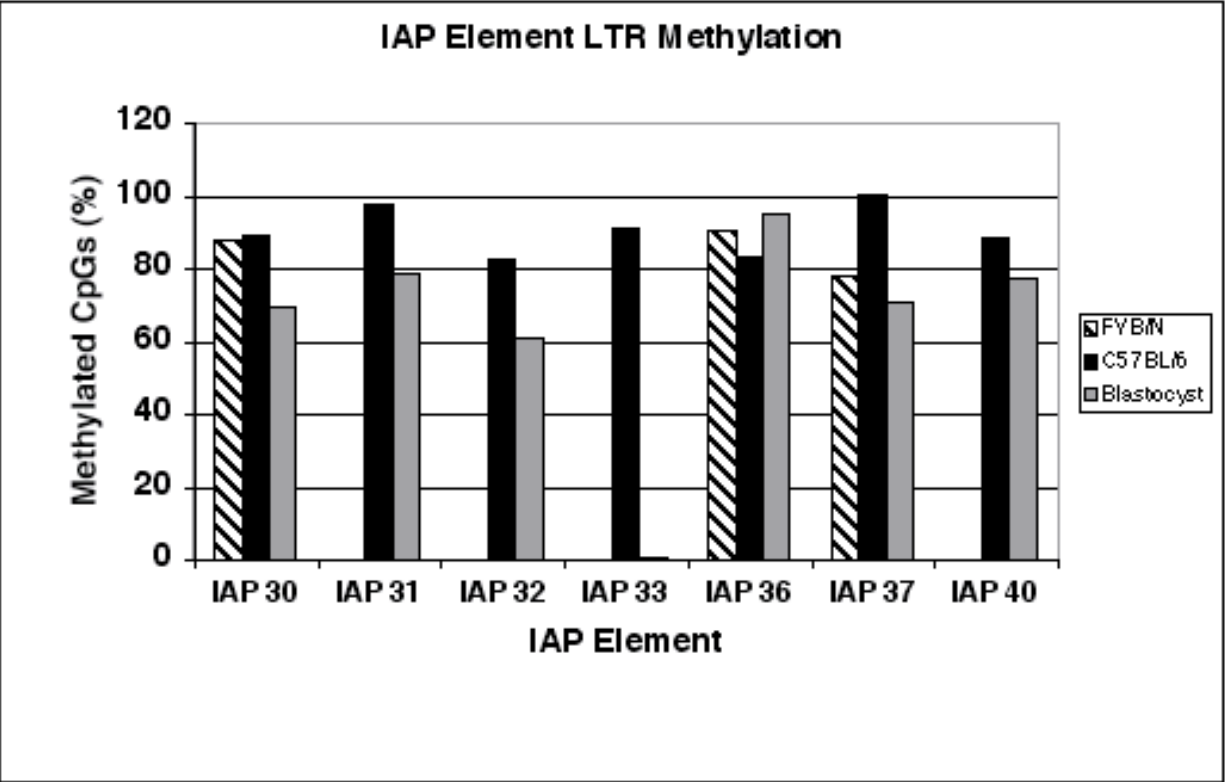


Figure 29. IAP elements are less methylated in the blastocyst than in the adult

IAP element LTRs were randomly selected from the mouse genome. The methylation of each IAP LTR was analyzed by bisulfite genomic sequencing. Bisulfite genomic sequencing was performed from a pool of 10 C57BL/6J blastocysts and from C57BL/6J adult genomic DNA. Certain IAP elements were also analyzed using FVB/N genomic DNA. The chart summarizes the bisulfite genomic sequencing data. The percentage of methylated CpG dinucleotides is shown on the y-axis. The data for each IAP element LTR in the FVB/N strain (hatched black bar), in the C57BL/6J strain (black bar), and in the blastocyst (gray bar) are presented along the x-axis. The percentage of methylated CpGs in each sample was calculated from the total number of methylated CpGs out of the total number of CpG dinucleotides sequenced.



Six of the seven randomly chosen IAP elements showed a high level of methylation in both blastocyst and adult tissues (Figure 29). Blastocyst methylation levels were always 12% to 20% lower than adult methylation levels. Loss of methylation in the blastocyst was due exclusively to the loss of one or a few methylated CpGs per IAP LTR. No IAP LTR sequences were completely unmethylated at the blastocyst stage. Little variation was seen in the methylation patterns of individual IAP LTRs in adult tissues for any of the six IAPs. Also, little difference was observed between the FVB/N and C57BL/6J DNAs. One exceptional IAP, IAP 33, was completely unmethylated in blastocyst DNA. No methylated CpGs were ever detected. In contrast, IAP 33 was methylated in adult tissues. One out of seven IAP LTRs selected from the mouse genome was completely unmethylated in blastocysts.

In summary, five undermethylated IAP elements were identified in the blastocyst. Three were identified in a library screen designed to preferentially amplify unmethylated IAP LTRs from blastocyst DNA. Two of these three IAPs were single LTR insertions. Correspondingly, one other single LTR insertion was chosen from the mouse genome and was found to be unmethylated at the blastocyst stage. The other unmethylated IAP was randomly selected from the mouse genome.

Table 10. Description of IAP element methylation and genomic location

The IAP elements identified in the blastocyst library screen (IAP 7, IAP 21, IAP 23, and IAP 29), and the IAP elements identified by BLAT search are listed. The approximate size of each IAP is described. The location of each IAP element near any known or predicted genes is described. All information was obtained using the UCSC genome browser (<http://genome.ucsc.edu/>). The results of methylation analyses for each IAP are summarized; the presence of > 75% methylated CpGs (Y), or the presence of < 25% methylated CpGs (N).

IAP Element	Size	Description	Methylation	
			Adult (Y/N)	Blastocyst (Y/N)
7	5.25 kb	5 kb upstream of <i>Galnt4</i>	Y	N
21	7.1 kb	250 kb from <i>Pmp22</i>	Y	Y
23	Single LTR	1 kb upstream of predicted gene (accession number AI355719)	Y	N
29	Single LTR	150 kb from <i>Foxd1</i> , within SGP predicted gene chr13_1516.1	Y	N
30	6.8 kb	75 kb upstream of <i>Shrm</i> , within predicted gene chr5_18.114	Y	N
31	7 kb	within <i>Rad51l-1</i>	Y	Y
32	7.2 kb	2.5 kb from <i>Mc2r</i>	Y	Y
33	7 kb	near no known or predicted genes	Y	N
35	Single LTR	8 kb downstream of <i>Titf1</i>	N	N
36	7 kb	within <i>Fmo1</i>	Y	Y
37	6.2 kb	300 kb from another IAP, 300 kb from <i>Slc8a1</i>	Y	Y
39	4.8 kb	40 kb downstream of <i>Snd1</i>	Y	Y
40	4.4 kb	Within reference sequence120000E11Rik	Y	Y

4.3.3. Comparison of unmethylated IAPs

The features of the five unmethylated IAPs were compared to find any similarities. The sequences of the 5' LTRs of the full length IAPs and the single LTRs were aligned (Figure 30). The alignment showed that all five IAPs have a typical LTR structure. The U3, R, and U5 regions were easily identified by comparison to published LTR sequences (MIARN IAP, Genbank accession number X01172) (Burt et al. 1984). The U3 regions and U5 regions of the IAP LTRs were well conserved with the exception of the IAP 23 LTR and the IAP 29 LTR. The IAP 23 LTR contained multiple differences in the 3' end of the U3 region, and the IAP 29 LTR contained several differences in the U5 region. Among all of the IAPs the majority of differences were seen in their R regions. This is expected for most IAPs.

The sizes and genomic locations of the IAPs were also compared (Table 10). IAP 7 is located on chromosome 10 and is 5.25 kb in size. The 5' LTR of IAP 7 is located 5 kb upstream of the 5' end of the *Galnt4* gene. IAP 23, IAP 29, and IAP 35 are single LTRs. IAP 23 is located on chromosome 13, approximately 1 kb from a predicted gene (accession number AL355719). IAP 29 is located on chromosome 13, 150 kb from the known gene *Foxd1* and within the SGP gene prediction chr13_1516.1 (predicted using mouse/human homology). IAP 35 is located on chromosome 12, approximately 8 kb downstream of *Titf1*. IAP 33 is 7 kb in size and is found on chromosome 3. However, IAP 33 is not located in the vicinity of any known or predicted genes in the database. These data show that the unmethylated IAPs are located on different chromosomes, and are of different classes based on their size. The locations of the IAPs near known or predicted genes was also very different. Four of the IAPs are located in or near known genes. However, many of the methylated IAPs are also located in or near known genes. These data do not provide a clear picture of a common feature that selects IAPs for demethylation.

Figure 30. DNA sequence alignment of unmethylated IAP element LTRs

Alignment of the IAP 7, IAP 33, IAP 23, IAP 29, and IAP 35 LTR sequences. Boxes separate the U3, R, and U5 regions of the LTR. The majority of gaps (.) in the sequence alignment occur in the R region of the IAP element LTRs. Sequence alignments were performed using Assemblylign software.

	U3
IAP 33	TGTTGGGAAGCCGCCCCACATTGCGCGTCACAAGATGGCGCTGACATCCTGTGTTCTAAGTTGGTAAACAAATAATCTGCGCATGAGCCAAGGGTAT.TTACGACTACTTGTACTCTGTTTTTCCCGTGAAACGTAGCTCGGCC.AT
IAP 7	TGTTGGGAAGCCGCCCCACATTGCGCGTTACAAGATGGCGCTGACATCCTGTGTTCTAAG.TGGTAAACAAATAATCTGCGCATGAGCCAAGGGTATCTTATGACTACTTGTGCTCTGCCTTCCCGTGA.CGTCAACTCGGCGGAT
IAP 35	TGTTGGGAAGCCGCCCCACATTGCGCGTCACAAGATGGCGCTGACATCCTGTGTTCTAAGTTGGTAAACAAATAATCTGCGCATGAGCCAAGGGTAT.TTACGACTACTTGTACTCTGTTTTTCCCGTGAAACGTAGCTCGGCC.AT
IAP 29	TGTAGGAAGCCGCGCTTACATTGCGCGTTGCAAGATGGCGCTGACATCCTGTGTTCTAAG.TGGTAAACAAATAATCTGCGCATGAGCCAAGGGTAGTTTTCCACCCCATGTGCTCTGCGTTCCCGTGA.CGACAACTCGGCGGAT
IAP 23	TGTTGGGAAGCCGCGCTTACATTGCGCGTTGCAAGATGGCGCTGACATCCTGTGTTCTAAG.TGGTAAACAAATAATCTGCGCATGAGCCAAGGGTAGTTCTTCACTCCATGTGCTCTGCCTTCCCGTGA.CGACAACTCAGCCGAT

	U3	R
IAP 33	GGGCTGCAGCCAATCAGGGAGTGATGCGTCTTAGGCAA.TTGTGTCTCTTTAAAGAGGAAAGGGTTT	GGTTTT.CTCTCTCTCTTGCTTCGCTCTCTTTGCTTCTTACACTCTGGCCCATAGATGTAAGCAATAAAGCT.T
IAP 7	GGGCTACAGCCAATCAGGGAGTGACACGTCGAGGCCAAAAGAGAATTCTCCTTAAAAAGGGACGGGGTTT	GGTTTT.CTCTCTCTCTTGCTTCTTGCTCTCTTGCTCTT
IAP 35	GGGCTGCAGCCAATCAGGGAGTGATGCGCGCTTAGGCAATGGTTGTCTCTTTAAATAGAAGGG..GTTT	GGTTTTCTCTCTCTCTTGCTTCCTCTC..TTGCTCTT
IAP 29	GGGCTGCAGCCAATCAGGGAGTGACACGTCCTAGGCGGAGGATAATTCTCCTTAAAAAGGGACGGGGTTT	GGCATT.CTCTCTCT..TGCT.....CTTGCTCTCTGCTCTGCTCTGAAGATGTAAGCAATAGAGCTCT
IAP 23	GGGCTGCAGCCAATCAGGGAGTGACACGTCCTAGGCGGAGGATAATTC.....	GGCATT.CTCTCTCT..TGCT.....CTTGCGCTCTGGC.....TCCTAAAGATGTAAGCAATAGAGCTCT

	R	U5
IAP 33	T.....	SCCGTA
IAP 7ACAC.....	SCCGCA
IAP 35ACACTCTGGCCGATAAAGATATAAGCAATAAAGCT.TT.....	SCCGTA
IAP 29TTCTCTTGCGCTCTGGCTCCTAAAGATATAAGCAATAGAGCTCTTGCTCTATCTCTCT..TGCTCCTAAAGATGTAAGCAATAAAGCTTT	SCCGCA
IAP 23	TGCTCTGGCTCTTGCACTCTTGCTCTTGCTCTTGCTCTTGAGCTCTTTTCGCTCTTGCTCTGGCTCCTAAAGATGTAAGCAATAGAGCTCTTGCTCTCTTGCACTCTTGCTCTGAAGATGTAAGCAATAAAG.TTT	SCCGCA

	U5
IAP 33	GAAGATTCTGGTT.GTTGTGTCTCTCTGGCGGTCGTGAGAACGCGTCAATAACA
IAP 7	GAAGATTCTGGTCTGTGTGTCTCTCTGGCGGTCGTGAGAACGCGTCAATAACA
IAP 35	GAAGATTCTGGT..GTTGTGTCTCTCTGGCGGTCGTGAGAACGCGTCAATAACA
IAP 29	GAAGATTCCGGTTTGTGTGTCTCTCTGGCTGGTCG.G.G.A.GCGTGTAAG
IAP 23	GAAGATTCTGGTTTGTTCGTTCTCTCTGGCGGTCGCGGCAACGCGTGAAGAGGA

4.4. Discussion

4.4.1. Single LTRs are unmethylated at the blastocyst stage

Interestingly, we identified two single LTRs in our blastocyst library screen (IAP 23 and IAP 29), both of which were located on chromosome 13. These single LTRs were unmethylated in the blastocyst and methylated in the adult (Figure 27). This suggested that single LTRs are preferentially demethylated at the blastocyst stage. In support of this notion a single LTR (IAP 35) identified using the UCSC genome database was also unmethylated at the blastocyst stage (Figure 28). These data suggest that one source of unmethylated IAPs in the blastocyst is a population of unmethylated, single IAP LTRs. Single LTRs are commonly found in the mouse genome. They are thought to be due to recombination between two LTRs, leaving a single LTR in the genome (Kuff and Leuders 1988). Unmethylated single LTRs may contain active promoters, and may affect neighboring gene transcription. However, they are not the sole source of unmethylated IAPs in the genome.

These data strongly support the idea that methylation at single LTR insertion sites is not properly maintained during preimplantation. Furthermore, at the IAP 35 LTR insertion site, the LTR was not targeted for *de novo* methylation during post-implantation development. IAP 35 was almost completely unmethylated in C57BL/6J genomic DNA. This difference in adult methylation may be attributed to the genomic context of the IAP element LTR. IAP 35 is located on chromosome 12, approximately 8 kb downstream of the gene *Titf1*. Several mouse ESTs and a genescan predicted gene are located within 5 kb of the IAP 35 LTR. *Titf1* is a homeobox transcription factor that is required for thyroid development (Perna et al. 1997). A requirement for active *Titf1* gene expression, or other nearby genes, may affect IAP 35 methylation levels in adult tissues.

4.4.2. A specific population of IAPs is unmethylated at the blastocyst stage

The demethylation of IAPs in the blastocyst could be sporadic or could target specific IAPs. Sporadic demethylation would be predicted to affect different IAPs in different blastocysts, and would lead to a mixed population of unmethylated and methylated LTRs in a pooled sample. The IAPs examined in these experiments, both methylated and unmethylated, showed consistent patterns of methylation. For example, every IAP 29 LTR was unmethylated in a pool of 10 blastocysts, and every IAP 36 LTR was methylated in the same sample (Figures 27 and 29). A consistent pattern of methylation on each LTR of an individual IAP should only be seen if that IAP is always unmethylated or always methylated. Therefore, our data suggest that specific IAPs are consistently demethylated at the blastocyst stage.

The presence of a specific group of unmethylated IAPs in the blastocyst indicates that a population of IAPs are able to escape methylation by the host. This is also observed in somatic tissues. Certain subsets of IAPs are specifically expressed in certain tissue types. For example, in mouse thymus tissue a limited group of related IAP elements are active (Meitz et al. 1992). Evidence indicates that this expressed population of IAP elements is hypomethylated. Examples of changes in IAP element methylation and expression are seen at specific IAP insertions in adult tissues. At the *nocturnin* locus a nearby IAP insertion is expressed in a rhythmic manner, similar to that seen for the *nocturnin* gene (Wang et al. 2001). The LTR of this IAP is demethylated due to its location near the *nocturnin* promoter. Also, in different tumor cell types, different subpopulations of IAPs are activated. This suggests that various changes in the host cell can lead to alterations in IAP element expression and presumably methylation (Dupressoir and Heidmann 1997). These examples suggest that certain IAPs are able to escape methylation in adult tissues as well as in the blastocyst. This is supported by the data obtained for IAP 35. IAP 35 is unmethylated in blastocyst DNA and in adult DNA.

Chapter 5: Analysis of trinucleotide repeat stability using the *RSVlgmyc* transgene

5.1. Introduction

Bacterial and yeast model systems, as well as human cell lines, and *in vitro* systems have been employed to study the characteristics of trinucleotide repeat expansion. Work in these areas has provided valuable information about trinucleotide repeat expansion. However, experiments performed in these systems do not address all of the features of repeat expansion as it occurs in humans. Unfortunately, attempts to create a mouse model system to study trinucleotide repeat expansion have not been successful. The reason behind the inability to create a mouse model system is still unclear.

5.1.1. Mouse models of trinucleotide repeat expansion

The *FMRI* locus is present in both the mouse and human genomes (Ashley et al. 1993). The *FMRI* proteins share 97% identity, and *FMRI*^{-/-} mice show phenotypes similar to those seen in fragile X patients (Dutch-Belgian Fragile X Consortium 1994). These data suggest that the two proteins perform similar functions. Interestingly, at the mouse locus the CGG repeat region is small, never exceeding nine CGG repeats. The small number of CGG repeats present at the mouse *FMRI* locus has not been observed to change in size. Several approaches have been taken to create a mouse model system to study the mechanism of CGG trinucleotide repeat expansion. Unfortunately, experiments using transgenic or knock-in techniques have not been able to closely model the characteristics of CGG repeat expansion. When size changes have been observed they are typically small, resulting in increases or decreases of only a few triplets.

Knock-in experiments were performed to address whether a large CGG repeat tract in the context of the mouse *FMRI* genomic locus would be targeted for expansion. A (CGG)₉₈ repeat

tract of human origin was introduced into the endogenous mouse *FMRI* locus in place of its small CGG repeat (Bontekoe et al. 2001). However, of 121 mice analyzed (80 maternal transmission and 41 paternal transmissions) only 15 CGG size changes were observed. The size changes seen were moderate changes, and the largest expansion documented was an increase of 10 CGG triplets. This suggests that the mouse genomic locus is not prone to the large repeat expansions seen at the human locus.

Many transgenic mouse lines have been generated that contain CGG repeat tracts of various sizes. A transgene containing a (CGG)₂₂ TGG (CGG)₄₃ TGG (CGG)₂₁ repeat was used to generate six transgenic mouse lines. In these lines a total of 342 animals were analyzed for up to four generations. However, each animal examined showed an identical repeat length (Lavedan et al. 1997). In subsequent experiments a series of transgenes were created that contained one of three CGG repeat tracts within the 5' UTR of the human *FMRI* locus, and subcloned upstream of a *LacZ* reporter. The human sequence included the first exon of the *FMRI* gene and the upstream CpG island. The CGG repeat tracts contained 32, 76, or 120 triplets in the organization (CGG)₉ AGG (CGG)₁₂ AGG (CGG)_x (Lavedan et al. 1998). The short and long CGG repeat tracts were stable in 151 transgenic mice analyzed.

Some transgenes have been generated that show moderate instability in transgenic animals. A YAC transgene containing the entire *FMRI* gene, including 300 kb of upstream sequence and 100 kb of downstream sequence, was used to generate transgenic animals (Peier and Nelson 2002). The transgene included a repeat tract of (CGG)₉ AGG (CGG)₉ AGG (CGG)₇₂. The resulting transgenic lines contained a range of sizes from 20 to 90 CGGs due to size changes that occurred in generation of the founder animals. After generation of founder animals only small size changes of -15 to +4 CGGs were observed. Thus, even though the entire

human locus was included in this transgene the parent-specificity and large size changes seen in human fragile X patients were never duplicated.

The most dramatic size changes documented for a CGG repeat containing transgene have come from studies done by Baskaran et al. in 2002. This transgene contained 1057 bp of human *FMR1* sequence, including a (CGG)₉ AGG (CGG)₆ AGG (CGG)₉ CGG repeat tract, downstream of an SV40 origin of replication. Repeat size changes in this transgene were analyzed in 95 mice for four generations. Again, repeat size changes occurred in the generation of founder animals. In the mouse line analyzed, the founder animal contained a repeat tract of approximately 196 repeats. This repeat tract showed further increases in size in the next generation to 279 repeats. However, the original small repeat tract was seen in all animals tested, even though the line was assumed to contain a single transgene copy. This may suggest that all of the size changes occurred in somatic tissues and were not due to germ line changes. Also, the repeat size changes were observed after both maternal and paternal transmission. Therefore, even though dynamic changes in size were seen, they do not model what is seen in human CGG repeat expansion. The authors suggest that the presence of the SV40 origin of replication is responsible for the size changes seen, possibly by providing a nucleosome free region of DNA or by providing a nearby origin of replication.

5.1.2. Establishing a transgenic mouse model for trinucleotide repeat expansion

The inheritance of fragile X syndrome has several important features that must be considered when attempting to model CGG repeat expansion. The CGG repeat expansion observed in fragile X syndrome is parent-specific, occurring only after passage through the female germ line. Also, full mutation CGG repeat tracts are methylated in patients carrying expanded alleles. The association between expanded, full mutation CGG repeat tracts and

methylation suggests that methylation may influence trinucleotide repeat stability. In order to test this hypothesis, we designed a transgene that should contain a CGG repeat tract with a high level of maternal-specific methylation.

In many ways the inheritance of trinucleotide repeat expansion in fragile X syndrome is similar to the inheritance of a maternal genomic imprint. At the *FMRI* locus and at a maternally imprinted locus the maternal allele is specifically targeted for modification, while the paternal allele is protected. In both examples the targeted alleles are subject to DNA methylation changes that are not seen on the opposite allele. Also, in each case the effect of the targeting process is an alteration in gene expression. One important distinction is that in genomic imprinting the targeting event occurs normally in the germ line and is necessary for normal development. However, in the mechanism of CGG repeat expansion the targeting event is abnormal, occurs infrequently, and leads to a heritable mutation.

It is possible that like a maternally methylated imprinted gene, the CGG repeat tract at the *FMRI* locus can be targeted for methylation specifically in the female germ line. Aberrant targeting of CGG repeats for methylation could affect repeat stability, leading to expanded, fully methylated alleles in some offspring. This mechanism would account for the parent-specificity of the process, due to the placement of a methylation mark in just one gamete. This mechanism also correlates with the presence of methylation on all full mutation, fragile X alleles in affected patients.

As described in Chapter 3, the *RSVlgmyc* transgene is a well characterized, mouse transgene that is imprinted at all sites of integration. *RSVlgmyc* is always maternally hypermethylated, and paternally undermethylated. Therefore, the *RSVlgmyc* transgene makes an excellent model system with which to test the effect of methylation on the stability of CGG

repeats in the mouse genome. The *RSVlgmyc* transgene has a well-defined differentially methylated domain (DMD). A CGG repeat tract placed in the vicinity of this DMD should also be specifically maternally methylated.

5.2. Transgene design

Due to the unstable nature of the CGG repeat region, it is subject to size changes during cloning in both bacterial and yeast vectors (Nichol and Pearson 2002). Consequently, the CGG repeat region used for this analysis was synthesized *in vitro*. Karen Usdin, a collaborator at the NIH, did the *in vitro* synthesis. The CGG repeat region was synthesized by a technique similar to that described in Lavedan et al. (1998). The repeat tract contains 120 total repeats, with 97 uninterrupted CGG repeats at its 3' end in the organization (CGG)₉ AGG (CGG)₁₂ AGG (CGG)₉₇. This organization is typical of the CGG repeat tracts found in fragile X patients. The length of this tract is within the premutation size range, and would be prone to expansion in humans. The CGG repeat was subcloned between *Mlu*I and *Nco*I sites in place of a 600 bp region of RSV and C α transgene sequence (Figure 31A). This region is adjacent to the 2.5 kb DMD of the original transgene and was predicted to be differentially methylated.

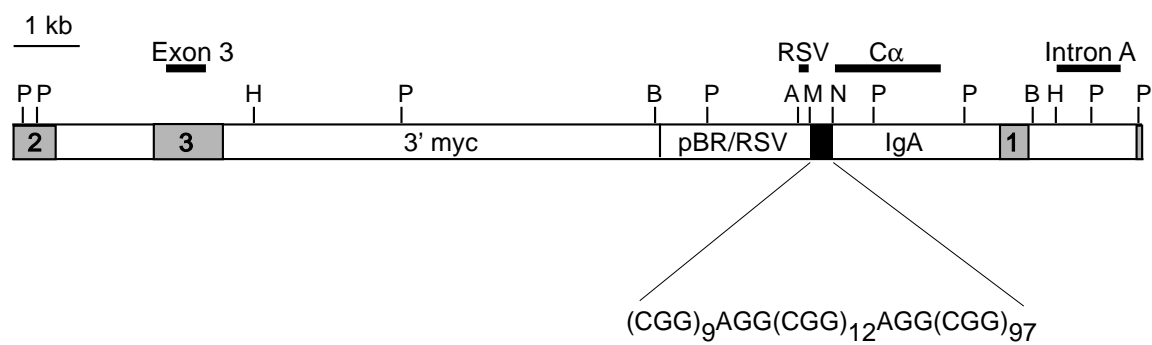
Figure 31. Generation of *CGG/Igmyc* transgenic mice

A, Schematic of the linear *CGG/Igmyc* transgene. The *CGG/Igmyc* transgene is a derivative of the *RSV/Igmyc* transgene and was generated in collaboration with Karen Usdin (NIH). *RSV/Igmyc* sequences are described in Figure 13A. Relevant restriction sites for Southern blots are indicated above the transgene: (P) *Pst*I, (H) *Hinc*II, (B) *Bgl*II, (A) *Apa*LI, (M) *Mlu*I, (N) *Nco*I. Probes for Southern blots are indicated as thick lines above the transgene. The CGG repeat tract is represented by a black box. The sequence of the CGG repeat tract is shown below the transgene.

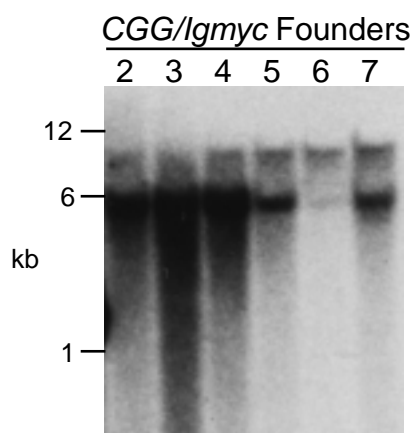
B, Southern blot of tail DNAs collected from transgenic founder animals. DNAs were digested with *Bgl*II and Southern blots were hybridized with the C \square probe. Lane numbers correlate with the number of the transgenic line (lines 2 through 7 are shown). DNA sizes indicated in kilobases (kb).

C, Southern blot of the DNAs described in panel B. DNAs were digested with *Pst*I and hybridized with the C \square probe. Lane numbers correlate with the number of the transgenic line (lines 1, 2, 3, 5, 6, and 7 are shown). Asterisks indicate the hybridization bands from the endogenous *Ig* locus (recognized by the C \square probe). DNA sizes are indicated in kilobases (kb). The 1.3 kb and the 2.5 kb bands are from the transgene locus. The 2.5 kb band contains the CGG repeat tract. Extra bands are predicted to be due to size changes in the repeat region.

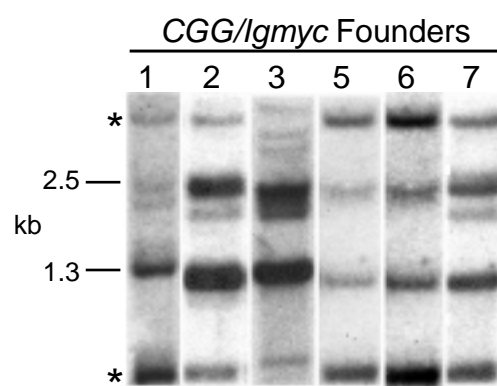
A



B



C



5.3. The CGG trinucleotide repeat showed instability following injection

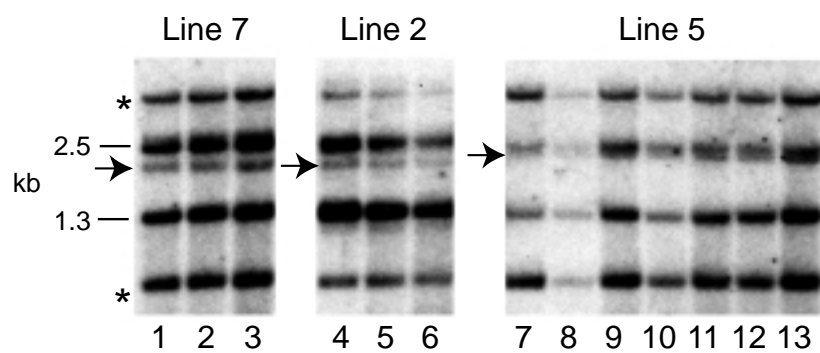
The linear *CGG/Igmyc* transgene construct was injected into the paternal pronucleus of FVB/N zygotes. Random integration of the transgene construct in the genome was tested by Southern blot analysis. Tail DNAs were collected from mice at the time of weaning (3 to 4 weeks). DNA from all possible founder mice was digested with the restriction endonuclease *Bgl*III, and Southern blots were hybridized with the C \square probe (Figure 31B). This digest and probe should recognize a 6 kb fragment specific to transgenic animals (Figure 31A). Seven transgenic founder mice were obtained and mated to wild type FVB/N mice to establish transgenic mouse lines.

The integrity of the CGG repeat tract was analyzed in founder animals by Southern blots. Genomic DNA was digested with the restriction endonuclease *Pst*I and Southern blots were hybridized with the C \square probe. This restriction digest and probe should produce two DNA fragments from the transgene, a 2.5 kb fragment that includes the repeat region, and an adjacent 1.3 kb fragment (Figure 31A). Two additional bands of 4.5 kb and 0.7 kb from the endogenous *Ig* locus should also be present. Interestingly, this Southern blot yielded different results in each transgenic animal tested (Figure 31C). In each animal the 1.3 kb DNA fragment was present, as was a DNA fragment of approximately 2.5 kb that should include the CGG repeat region. However, along with the predicted transgene DNA fragments, DNA fragments of slightly different sizes were also obtained in each founder animal (indicated by arrows, Figure 31C). These bands migrated close to the CGG repeat containing band, and suggested that size changes occurred in the CGG repeat region in one or more copies of the transgene.

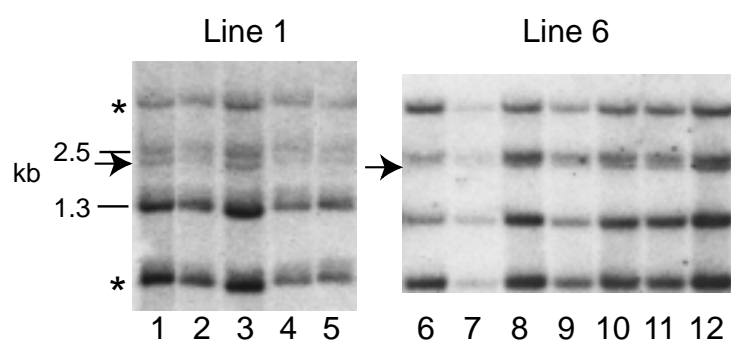
Figure 32. Analysis of trinucleotide repeat stability in *CGG/Igmyc* transgenic lines

Founder animals carrying the *CGG/Igmyc* transgene were crossed to wild type FVB/N mice and transgenic lines were established. Each panel includes representative Southern blots of DNA samples collected from the founders for each line and their progeny. DNAs were digested with *Pst*I and hybridized with the C \square probe. Asterisks indicate the hybridization bands from the endogenous *Ig* locus. Extra bands (indicated by arrows) are predicted to be due to size changes in the repeat region. DNA sizes are indicated in kilobases (kb). A, Data from transgenic line 2 (lanes 4-5), transgenic line 5 (lanes 7-13), and transgenic line 7 (lanes 1-3) are shown. Founder animals for each line are shown in lanes 1, 4, and 7. B, Data from transgenic line 1 (lanes 1-5) and transgenic line 6 (lanes 6-12) are shown. The founder animals for each line are shown in lanes 1 and 6. C, Data for transgenic line 3 are shown. The founder female is shown in lane 1. This Southern blot includes wild type FVB/N DNA (lane 9). This lane demonstrates that the two bands indicated by asterisks are from the endogenous *Ig* locus.

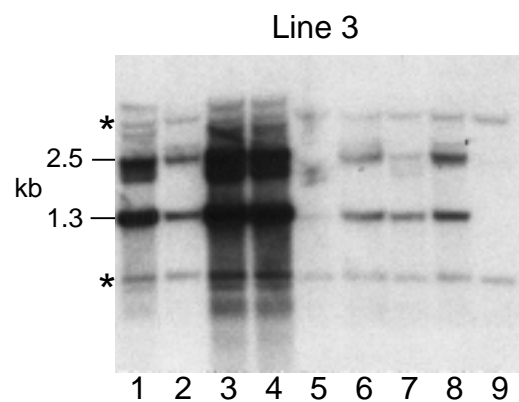
A



B



C



Founder animals were mated to wild type FVB/N mice to establish transgenic lines. Six of the seven founder animals were fertile and passed the transgene to their offspring. Five of the transgenic lines contained unique insertion sites, inherited by all future progeny (Figure 32A and 32B, lines 1, 2, 5, 6, 7). All of the transgenic progeny of a founder animal contained identical band patterns by Southern blot analysis of *Pst*I digested DNAs probed with C \square . However, one transgenic founder animal (line 3) contained multiple insertion sites that assorted independently in the founder female's offspring. This resulted in more than two distinct band patterns by Southern blot analysis of *Pst*I digested DNAs that were slightly different than the pattern seen in the founder animal (Figure 32C). The presence of multiple insertion sites following transgene injection has been described by others (Tomoko et al. 2002). Line 3 was subdivided into 2 lines, *CGG/Igmyc-3-1* and *CGG/Igmyc-3-2*, which were characterized further.

The *CGG/Igmyc-3-1* and *CGG/Igmyc-3-2* lines were maintained by mating to wild type FVB/N mice (Figure 33A). Progeny from both lines yielded consistent patterns of bands by Southern blot analysis, indicating that they represent unique insertion sites isolated from the original mosaic founder animal (Figure 33B and 33C). This was confirmed by Southern blot analysis using restriction enzymes and probes at each end of the transgene construct (Figure 33D). The restriction enzyme *Bgl*II has one recognition site at the end of the construct and one recognition site in an unknown region in the genomic flanking DNA. The intron A probe at the end of the construct should pick up different bands at each unique insertion site. The *Hinc*II restriction enzyme and the exon 3 probe were used to examine the opposite end of the construct. It is clear from the different patterns of bands found in the *CGG/Igmyc-3-1* and *CGG/Igmyc-3-2* lines with each Southern blot that they represent two different insertion sites.

Figure 33. Multiple transgene insertion sites in the *CGG/Igmyc-3* transgenic line

A, Pedigree for the *CGG/Igmyc-3* transgenic line. The founder female was mated to an FVB/N male to establish a transgenic line. Unfilled circles, wild type females; unfilled squares, wild type males; partially filled circles, hemizygous transgenic females; partially filled squares, hemizygous transgenic males. Numbers correspond to the lane numbers in panels B and C. B, Southern blot of DNA samples collected from carriers of the transgene in the *CGG/Igmyc-3* transgenic line. DNAs were digested with *Pst*I and hybridized with the C \square probe. Asterisks indicate the hybridization bands from the endogenous *Ig* locus. DNA sizes are indicated in kilobases (kb). A male carrier of the transgene (lane 1) was mated and his progeny are shown in lanes 2-9. C, Southern blots performed as described in panel B. DNA samples were from a male carrier of the transgene (lane 10) and his progeny (lanes 11-17). D, Southern blots performed on the DNA samples from lane 1 in panel B (lanes 1 and 3) and lane 10 in panel B (lanes 2 and 4). Lanes 1 and 2 show DNAs digested with *Bgl*II and hybridized with the intron A probe. Lanes 3 and 4 show DNAs digested with *Hinc*II and hybridized with the exon 3 probe. DNA sizes are indicated in kilobases (kb).

In order to characterize the nature of the size changes seen in generation of the founder animals, and to interpret any future changes in repeat tract length, it was important to determine the number of transgene copies present at each insertion site. A low copy number is preferred, as a large number of transgene copies will provide a large number of CGG repeat regions to account for during analysis. The copy number of each transgenic line was estimated by Southern blot analysis of genomic tail DNAs digested with the restriction endonuclease *Pst*I, and hybridized with the C \square probe. As described above, a *Pst*I digest yields distinct DNA fragments from the transgene, and from the endogenous *Ig* locus. By comparing the intensity of the endogenous bands (two copies per diploid mouse genome), to the intensity of the transgene bands, an approximation of transgene copy number was possible. In three transgenic lines the transgene was present in greater than three copies (*CGG/Igmyc-2*, *CGG/Igmyc-5*, and *CGG/Igmyc-7*) (Figure 32A). In four lines the transgene was present in one or a few copies (*CGG/Igmyc-1*, *CGG/Igmyc-3-1*, *CGG/Igmyc-3-2*, *CGG/Igmyc-6*) (Figure 32B and 32C). The low copy number transgenes were chosen for subsequent analysis.

5.4. Size changes occurred within the repeat region

5.4.1. Southern blot analysis

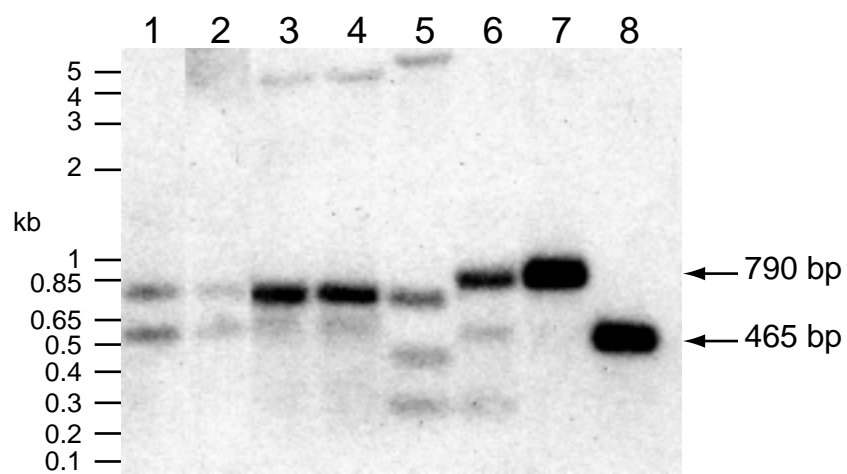
Additional Southern blots were performed to determine if the predicted size changes occur within the CGG repeat tract. DNA samples from each transgenic line were digested with *Apa*LI and *Nco*I and Southern blots were hybridized with the RSV probe. The *Nco*I site is located at the 3' end of the repeat region, and the *Apa*LI site is located 160 bps away from the CGG repeat region. This digest isolates the CGG repeat tract and 160 bps of the adjacent RSV sequence (Figure 31A). Based on the size of the original CGG repeat tract the RSV probe should recognize a 520 bp band on the Southern blot. The *Apa*LI plus *Nco*I digest includes only 160

bps of RSV sequence, and any size changes observed most likely occurred within the CGG repeat region.

To accurately determine the size of the repeat regions in the *CGG/Igmyc* transgenics, DNA from a carrier of the *RSV/Igmyc* transgene was digested with *Apa*LI plus *Nco*I or *Apa*LI plus *Eco*RI to yield fragments of 790 bp and 465 bp respectively, and these samples were run in parallel with the *CGG/Igmyc* genomic DNAs (Figure 34). For each transgenic line the CGG repeat tract size changes predicted from the *Pst*I digest are also observed after the *Apa*LI plus *Nco*I digest. This suggests that the bands seen by both Southern blots represent changes in the size of the repeat tract. The size changes seen were predominantly contractions, with the exception of one expansion. The result was a variety of band sizes in each transgenic line due to the multicopy nature of transgene insertions. These data also demonstrated that the predominant CGG repeat containing DNA fragment migrates near the 750 bp band on an agarose gel. The 750 bp band represents the starting size of the repeat tract (120 triplets). This was demonstrated by Southern blots of genomic DNAs run in parallel to the injected transgene construct (data not shown).

Figure 34. The repeat tract changed in size upon generation of founder animals

Southern blot of DNAs from *CGG/Igmyc* transgenic animals from line 6 (lanes 1 and 2), line 3-2 (lanes 3 and 4), line 3-1 (lane 5), and line 1 (lane 6). DNAs were digested with *Apa*LI and *Nco*I and run on a 1.2% agarose gel. The Southern blot was hybridized with the RSV probe. DNA fragments of known sizes were included in the Southern blot as a size standard. Lane 7 contains DNA from a *RSV/Igmyc* transgenic animal digested with *Apa*LI and *Nco*I (790 bp band indicated by an arrow). Lane 8 contains DNA from a *RSV/Igmyc* transgenic animal digested with *Eco*RI and *Nco*I (465 bp band indicated by an arrow). DNA ladder sizes are shown in kilobases (kb).



The *Apa*LI plus *Nco*I Southern blots were used to approximate the number of CGG triplets lost or gained in each transgenic line based on the size of the initial CGG repeat tract. In the *CGG/Igmyc-1* transgenic line the size changes were the most dramatic (Figure 34, lane 6). The band containing the original repeat tract was absent. The repeat tract increased in size by approximately 20 to 30 CGG triplets in at least one transgene copy (820 bp band), and decreased in size by approximately 40 to 50 CGG triplets in another transgene copy (600 bp band). In the *CGG/Igmyc-3-1* transgenic line the original repeat tract was present in at least one copy (750 bp band), and decreased by approximately 80 to 90 repeats in another copy (480 bp band) (Figure 34, lane 5). A small band that appeared to represent almost a complete loss of the repeat tract was also seen (350 bp band). In the *CGG/Igmyc-3-2* transgenic line the original repeat tract was present in at least one copy (750 bp band), and decreased by approximately 50 to 60 repeats in another copy (600 bp band) (Figure 34, lanes 3 and 4). Similar results were seen for the *CGG/Igmyc-6* line (Figure 34, lanes 1 and 2), the original repeat size is seen and a band that represents loss of approximately 65 to 70 triplets (600 bp band). The approximate size changes in the CGG repeat tract are summarized in Table 11.

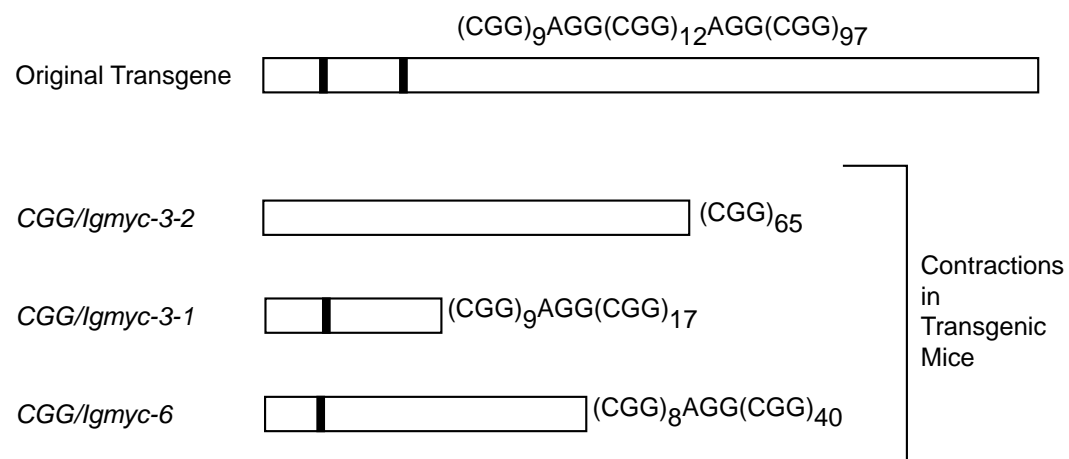
Table 11. CGG repeat expansions and contractions in *CGG/Igmyc* transgenic lines

For each transgenic line the number of CGG repeats (number of triplets) is an approximation based on Southern blot data (*Apa*LI + *Nco*I restriction digest and RSV probe). The number of transgene copies per diploid genome was estimated by comparing the intensity of the transgene hybridization bands to the intensity of the hybridization bands from the endogenous *Immunoglobulin* heavy chain locus on a Southern blot hybridized with the C α probe.

Transgenic line	Copy number	Repeat tract length (number of triplets)	Generations
<i>CGG/Igmyc-1</i>	2	150/ 70	8
<i>CGG/Igmyc-3-1</i>	3	120/ 40/ 0	7
<i>CGG/Igmyc-3-2</i>	2	120/ 70	9
<i>CGG/Igmyc-6</i>	2	120/ 50	8

Figure 35. Sequencing of the CGG repeat tract contractions

Summary of results obtained from sequencing the CGG repeat tracts from three transgenic lines (*CGG/Igmyc-3-2*, *CGG/Igmyc-3-1*, *CGG/Igmyc-6*). White rectangles indicate the CGG repeat region. Black bars indicate AGG interruptions. The top line represents the original repeat size; the sequence of the repeat is shown above. The bottom three lines depict the contractions seen in each transgenic lines; actual sequences are shown to the right.



5.4.2. Direct sequencing of the repeat region

To confirm that the size changes occurred in the repeat region of the transgene, the repeat region was directly sequenced. Sequencing of the repeat region was performed for three lines. Genomic DNA was PCR amplified with primers flanking the repeat tract. The PCR products were sequenced with a primer immediately adjacent to the 5' end of the repeat tract. Direct sequencing from the PCR products was performed to avoid size changes upon subcloning the CGG repeats into bacterial vectors. Only contractions of the repeat tract were sequenced because small CGG repeat tracts were preferentially amplified in the PCR reaction. The sequencing results are summarized in Figure 35.

The original repeat tract was composed of 120 triplets with AGG interruptions at positions 10 and 23 ((CGG)₉AGG(CGG)₁₂AGG(CGG)₉₇). The 5' of the repeat tract in each sequenced PCR product was intact. In line 3-2 the first AGG interruption was absent and the repeat tract consisted of 65 pure CGG triplets. In line 3-1 the first AGG interruption was present, and the second AGG interruption was absent. Following the first AGG interruption the repeat tract was composed of 17 pure CGGs. In line 6 the first AGG interruption was present, however it was shifted by 1 position. The 3' end of the repeat tract contained 40 pure CGG triplets with no second AGG interruption. For each sample the sequence confirmed that the changes occurred at the 3' end of the repeat. These results demonstrate that the size changes seen by Southern blot occurred within the repeat tract. The approximations made by comparing the band size of the initial repeat, to the band sizes of repeat tract contractions and expansions by Southern blot were accurate to within 10 to 15 CGG triplets. These data confirm that Southern blots are a relatively accurate method to determine the number of CGG triplets gained or lost.

5.5. The CGG repeat is stable in somatic cells

The CGG repeats found in human fragile X patients often show mitotic instability, resulting in individuals with different repeat tract lengths in different somatic tissues. Southern blots were performed on DNA samples isolated from the heart, liver, kidney, and spleen of the *CGG/Igmyc* transgenic animals to identify any size changes in somatic tissues. Tissue DNA samples were digested with *Pst*I and Southern blots were hybridized with the C \square probe. Results are shown for the *CGG/Igmyc3-2* transgenic line (Figure 36). No obvious differences were seen in the size of the 2.5 kb repeat containing band in any tissue examined. Similar results were obtained for the *CGG/Igmyc3-1* and *CGG/Igmyc-1* transgenic lines (Data not shown).

5.6. The transgene showed intergeneration stability

The stability of the CGG repeat tracts present within each transgenic line was analyzed over seven generations in all four transgenic lines (Table 11). In each generation the transgene was inherited through either the male or female germ line. Transgenic animals were analyzed for changes in repeat size by Southern blots with the *Apa*LI plus *Nco*I restriction digest and RSV probe. A representative Southern blot of each transgenic line is shown (Figure 37). No changes in repeat size were observed. It is possible that small changes in repeat size have gone unnoticed due to the limits of detection possible by Southern blot analysis. However, we are interested in obtaining large changes in repeat size, and the loss or gain of a few repeats is not of great interest.

Figure 36. The CGG repeat is stable in somatic tissues

Tissue DNAs were collected from one transgenic carrier of the *CGG/Igmyc-3-2* transgenic line. The Southern blot shows tissue DNAs digested with *Pst*I and probed with C \square . DNA sizes are indicated in kilobases (kb). Asterisks indicate hybridization bands from the endogenous *Ig* locus.

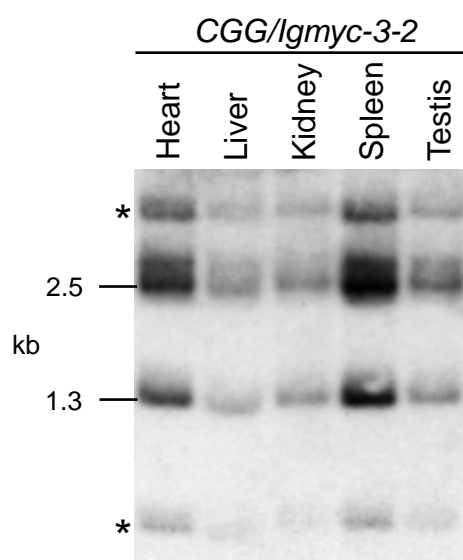
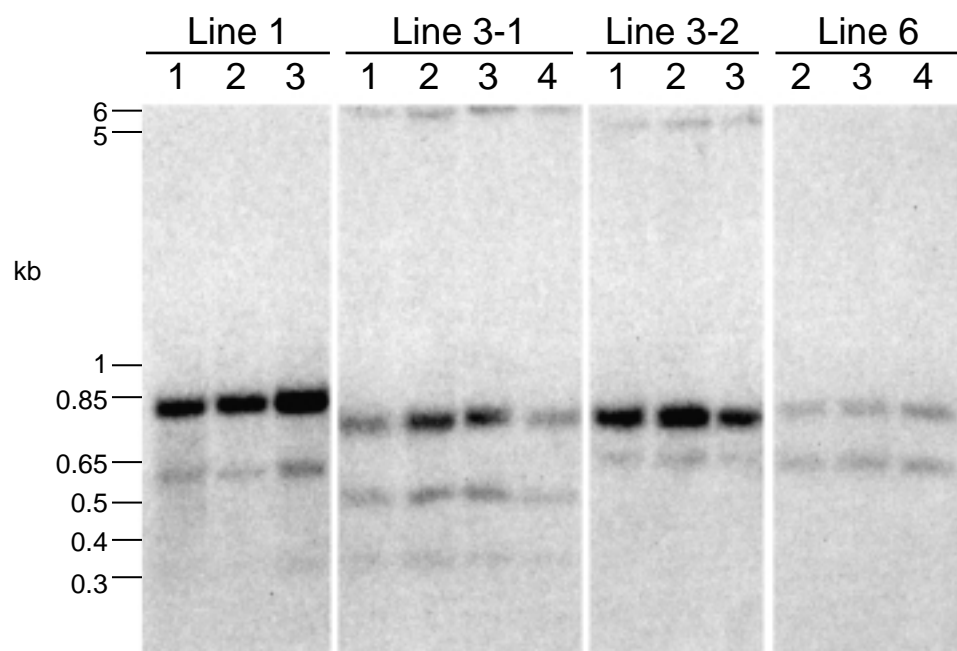


Figure 37. The CGG repeat is stable over successive generations

Southern blot of DNAs from *CGG/Igmyc* transgenic animals from line 1, line 3-1, line 3-2, and line 6. DNA samples are from representative animals over successive generations (the generation number is indicated at the top of each lane). DNAs were digested with *Apa*LI and *Nco*I and the Southern blot was hybridized with the RSV probe. DNA sizes are indicated kilobases (kb).



5.7. Maternal methylation of the transgene is not consistent

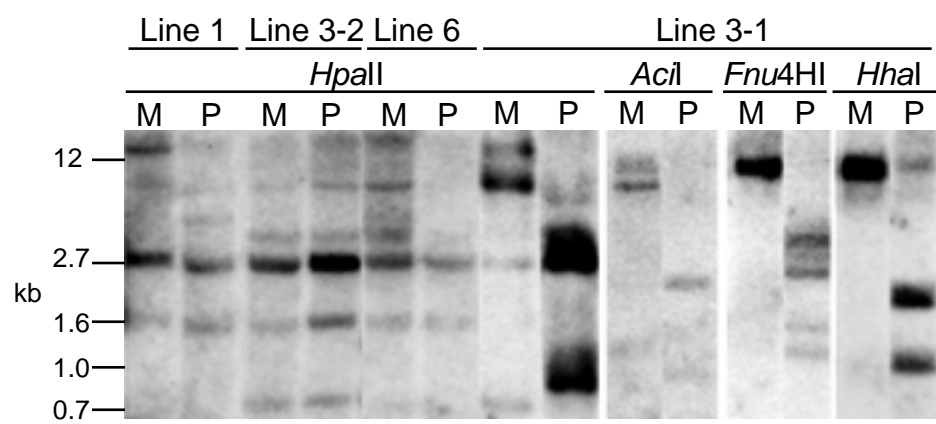
The *CGG/Igmyc* transgene was designed to test the effect of maternal-specific methylation on the stability of the CGG repeat region. The CGG repeat tract was positioned adjacent to the DMD of the original *RSV Igmyc* transgene, and minimal DMD sequences were removed. The methylation of maternal and paternal alleles in each transgenic line was compared by Southern blot. DNA samples were collected from hemizygous carriers of a maternally or paternally inherited transgene, and cleaved with the methylation sensitive restriction endonuclease *HpaII* (Figure 38). Southern blots were hybridized with the C \square probe.

In transgenic line 1 the maternal alleles were slightly more methylated than the paternal alleles. In transgenic line 3-2 the maternal and paternal alleles showed equivalent levels of methylation. In transgenic line 6 the maternal alleles were slightly more methylated than the paternal alleles. Finally, in transgenic line 3-1 the maternal alleles were highly methylated and the paternal alleles were undermethylated. However, in all three lines that showed differential methylation, the methylation patterns were not consistently observed in each mouse analyzed.

The most distinct differential methylation observed occurred in transgenic line 3-1. This was confirmed with the *HhaI*, *Fnu4HI* and *AciI* methylation sensitive restriction enzymes (recognition sequences GCGC, GCNGC, and GGCG respectively). The *Fnu4HI* and *AciI* enzymes recognize CpG sites within the CGG repeat tract itself. Maternal alleles that showed a high level of methylation at *HpaII* sites also showed a high level of methylation at sites within the repeat tract. Likewise, paternal alleles that were undermethylated at *HpaII* sites were also undermethylated at CpG sites within the repeat tract. These data suggest that the transgene is capable of establishing differential methylation at the CGG repeat tract. However, even in mice that showed maternal-specific methylation the repeat was stable.

Figure 38. The *CGG/Igmyc* transgene is inconsistently differentially methylated

Southern blot analysis of genomic tail DNAs digested with the methylation sensitive restriction endonuclease *HpaII* and hybridized with the C \square probe. Hemizygous carriers of the *CGG/Igmyc* transgene from transgenic lines 1, 3-2, 6, and 3-1 are shown. Southern blots were also performed using the methylation sensitive restriction endonucleases *AciI*, *Fnu4HI* and *HhaI* on the DNAs from transgenic line 3-1. DNAs were hybridized with the C \square probe. (M) Maternal inheritance. (P) Paternal inheritance. DNA sizes are indicated in kilobases (kb).



5.8. Discussion

5.8.1. Differences between repeat expansion in the human and mouse

Many unsuccessful attempts have been made to model CGG repeat expansion in the mouse. No model to date has been able to reproduce the dynamic instability of the CGG repeat in fragile X syndrome. This is also the case with many other trinucleotide repeats. No mouse models to date can completely recapitulate the features of trinucleotide repeat expansions seen at disease loci. The *HD* and *DRPLA* loci both contain CAG repeat tracts, and the *DMPK* locus contains a CTG repeat (described in Table 2). At these loci trinucleotide repeat expansions lead to Huntington disease, dentatorubral-pallidoluysian atrophy, or myotonic dystrophy type I respectively. Recent attempts have been moderately successful at generating instability in these trinucleotide repeat tracts.

Transgenic mice have been generated that contain unstable CAG repeats (Mangiarini et al. 1997; Sato et al. 1999). In one model a 1.9 kb region from the 5' end of the *HD* gene was used as a transgene. The CAG repeat contained 116, 141, or 144 repeats depending upon the transgenic line. All three lines showed age-dependent, somatic instability of the repeat region. In one line, paternal transmission resulted in instability. The largest size change observed in the repeat tract was an expansion of 10 CAG triplets. In another line, a consistent tendency was observed for paternal increases and maternal decreases in repeat size. Similarly, a transgene containing the entire human *DRPLA* genomic sequence with 78 CAG repeats was used to generate three mouse lines (containing 76, 77, or 78 repeats). In one line paternal transmission led to expansions and contractions ranging between -2 to +1 repeats, and maternal transmission led to contractions of -1 to -3 repeats. In this mouse line the mutation rate increased with age, and somatic mosaicism was seen in carrier animals.

These results suggest that it is possible to generate mouse models of trinucleotide repeat instability. However, unlike CGG repeats, CAG repeats in humans do not normally show large changes in repeat length (La Spada et al. 1994). At the Huntington's disease locus premutation alleles range from 36 to 39 CAGs with small expansions of 1-4 repeats, and large expansions of greater than 7 repeats. These expansions occur primarily through the paternal germ line and patients show marked somatic mosaicism. Likewise, dentatorubral-pallidoluysian atrophy (DRPLA) is caused by expansion of a CAG repeat by an average of 5 triplets, primarily after paternal transmission.

A CTG repeat at the myotonic dystrophy (DM) locus ranges from 5 to 37 repeats in the population and can expand to 50 to 1000 repeats in affected patients. The repeat shows somatic instability and parent-specific expansion. These features are similar to the features of CGG repeat expansion. Variable results have been obtained with mouse models of CTG repeat expansion. A transgene containing 162 CTGs showed size changes of -7 to +2 after maternal transmission and size changes of -11 to +7 after paternal transmission (Monckton et al. 1997). Similarly, transgenes containing 45 kb of human genomic sequence with a 55 CTG repeat showed moderate size changes of -1 to +6 (Gourdon et al. 1997). However, these moderate size changes are not typically seen in DM patients.

Recently, a similar transgene, containing 45 kb of human genomic sequence, was modified to include over 300 CTGs (Seznec et al. 2000). Different founder lines contained slightly different repeat numbers, 304, 362, or 362/184/147 in a multicopy transgene. Expansions were documented in these lines of up to 60 CTGs in one generation with a paternal bias for repeat expansion. These lines also showed increased repeat expansion with increased

age of the transmitting male, and large differences in somatic tissues. These features more closely model what is expected of the CTG repeat in human DM patients.

Importantly, these data demonstrate that it is possible to recreate the dynamic changes in repeat size in a mouse model. However, these data also point out relevant differences between the mouse and human systems. The CTG repeat tract did not expand at 55 CTGs, which is in the premutation range for CTG repeat expansion. However, the repeat did expand at 300 CTGs, well beyond the number of repeats needed to see repeat expansion in humans. These results may demonstrate that the threshold for expansion in mice is larger than in humans. The CGG repeat tract in the *CGG/Igmyc* transgene contained 120 triplets, well within the premutation range seen in humans. However, based on the above results it is possible that increasing the size of the CGG repeat tract in the mouse model may increase its tendency to expand.

Also, some successful mouse model systems have employed a large region of genomic sequence from human disease loci to trigger expansion. This suggests that unique features of the human genomic sequence, which are not found in the mouse genome, are required to trigger expansion. A human-specific aspect of repeat expansion is suggested by the fact that large trinucleotide repeats are not observed in the mouse. The CGG repeat region at the mouse *FMRI* locus only contains around 9 CGG repeats. Experiments replacing the small mouse repeat with a larger repeat in the context of the mouse *FMRI* locus did not trigger repeat expansion (Bontekoe et al. 2001). However, including a large amount of human genomic sequence cannot be the only factor affecting repeat expansion in the mouse. A YAC transgene containing 400 kb of genomic DNA surrounding the human *FMRI* locus was not unstable in transgenic animals (Peier and Nelson 2002).

5.8.2. DNA repair and DNA replication

Trinucleotide repeats are unstable in both yeast and bacteria (Hirst and White 1998). Similar to what is seen in humans, repeats in these systems show length dependent expansions and contractions. Experiments in *E. coli* have elucidated that DNA replication and DNA repair affect the stability of trinucleotide repeats (Balakumaran et al. 2000). For example, mutations affecting double strand break repair proteins increase the stability of the repeat. The orientation of the repeat tract relative to the origin of DNA replication also affects the stability of trinucleotide repeats. This is observed as an increase in the occurrence of contractions and expansions when the repeat is in one specific orientation (Hirst and White 1998). It has also been shown that *E. coli* strains carrying mutations in mismatch repair enzymes, nucleotide excision repair enzymes, Okazaki fragment processing enzymes, and DNA polymerase III show variations in the number of expansions and contractions seen compared to wild type strains (Iyer et al. 2000).

Similar studies in the yeast *Saccharomyces cerevisiae* have shown that certain mutations in DNA repair proteins, DNA replication proteins, and the orientation of the repeat with respect to an origin of replication can affect repeat stability (White et al. 1999). Mutations in homologues of the human BLM and WRN helicases decrease contractions. Also, a mutation in Rad27p, a protein involved in processing Okazaki fragments during DNA replication, increases the rate of expansions. The involvement of DNA replication in the process of repeat expansion in this system is suggested by the orientation dependent expansions of 5 to 40 CGG repeats observed in yeast.

The above observations in the bacterial and yeast model systems suggest that changes in DNA replication and the fidelity of DNA repair alter the stability of trinucleotide repeats. The effect of mutations in DNA repair proteins has also been observed in the mouse. Expansion of

CTG repeats in the transgene containing 45 kb of human genomic sequence and 300 CTGs was altered in various mutant mouse strains (Savouret et al. 2003). For example an *Msh2* mutation (a mismatch repair protein) drove the mutability of the repeat from expansions to contractions. A mutation in another DNA repair protein, Rad52, decreased the size of expansions observed.

The ability of alterations in the DNA repair and replication processes to increase or decrease the expansions and contractions seen at trinucleotide repeat tracts suggests that these processes are involved in the mechanism of repeat expansion. Many investigators have suggested that the formation of slipped-strand structures during DNA replication (S-DNA) is involved in the expansion of trinucleotide repeats. In support of this, it was shown that the formation of slipped-strand structures increases in long pure CGG repeat tracts and decreases in repeat tracts with AGG interruptions (Pearson and Sinden 1998).

These data suggest that mating the *CGG/Igmyc* transgene into certain mutant mouse backgrounds may influence the stability of the repeat. The stability of CGG repeats in mice, compared to their instability in humans, suggests that there are subtle differences in the DNA replication and repair processes between the mouse and human systems. Determining the specific mutations in the mouse that trigger repeat expansion may clarify what process ultimately leads to repeat expansion in humans. For instance, mutations in the Rad27p protein in yeast increase the rate of repeat expansions. This protein is involved in processing of Okazaki fragments during DNA replication. The formation of slipped strand structures during DNA replication across a CGG repeat region may interfere with the function of this protein and lead to expansion. Therefore, mutations in the mouse homologue of Rad27p, FEN1, that alter its function may lead to repeat instability in mice.

5.8.3. Size changes and transgene injections

In many trinucleotide repeat transgenic mouse lines the repeat tracts have increased or decreased in size following injection into the mouse zygote (Peier and Nelson 2002; Baskaran et al. 2002). However, upon establishment of a mouse line the repeats showed no germ line instability. This is similar to what we have observed in generation of *CGG/Igmyc* transgenic animals. Each founder animal examined showed size changes of the injected DNA construct. However, these size changes remained stable in the established mouse line. This indicates that the CGG repeat is stabilized following integration into the mouse genome. The nature of the instability seen in generation of transgenic animals is not clear.

5.8.4. Methylation and repeat stability

The original intention of these experiments was to investigate the effect of DNA methylation on the stability of the trinucleotide repeat. However, in the transgenic lines that were established the repeat region was not consistently differentially methylated. The effect of methylation on repeat stability in a mouse model system is still an important question to address. Recent experiments in a bacterial model system suggest that methylation of the CGG repeat may stabilize the repeat (Nichol and Pearson 2003). In this system a pure repeat of 53 CGGs or an interrupted repeat with 32 pure CGGs at the 3' end was stabilized by methylation. Similarly in a COS1 primate cell line premethylation of a trinucleotide repeat decreased the occurrence of contractions following transfection (Nichol and Pearson 2002).

It may be possible to design a different transgene to get at this question. Perhaps by moving the CGG repeat within the transgene DMD, or removing less of the DMD when introducing the CGG repeat tract. However, these approaches do not guarantee that the CGG region will be methylated. It may also be useful to methylate the CGG repeat *in vitro* prior to

injecting the construct. The instability observed in each transgenic consistently occurred upon injection of the construct. If methylation increases or decreases repeat stability methylating the transgene prior to injection, and analyzing the stability of the repeat in founder animals should give a clear result. Overall, these experiments have confirmed the results obtained by many other investigators; trinucleotide repeats are inherently stable in the mouse genome.

Chapter 6: Summary and Future Directions

The experiments described above were initiated in an attempt to gain a better understanding of DNA methylation in mammalian species. Establishing and maintaining proper DNA methylation is critical for normal mouse development. This was clearly illustrated by the embryonic lethality observed with the loss of the Dnmt1 methyltransferase (Li et al. 1992). The expression of imprinted genes is altered following loss of methylation, and the expression of IAP elements is dramatically increased (Li et al. 1993; Walsh et al. 1998). For both imprinted genes and IAP elements, methylation is established in the gamete and maintained during preimplantation development (Tremblay et al. 1997; Lane et al. 2003). For imprinted genes methylation is parent-specific, and for IAPs methylation is present on both parental alleles. These are exceptional sequences, as the bulk of the mouse genome (including a subset of IAP sequences) loses methylation at this time. This suggests that these sequences possess specific features that allow for methylation maintenance during preimplantation development.

Interestingly, it has been shown that loss of the Dnmt1o methyltransferase during preimplantation development has an effect on imprinted gene methylation and expression (described in section 1.5.5). The *Dnmt1^o* mutation eliminates the Dnmt1o protein from oocytes and preimplantation embryos in homozygous *Dnmt1^o* females (Howell et al. 2001). Embryos do not develop normally and die in the last third of gestation with variable phenotypes. Experiments examining the methylation of these embryos have clearly shown that methylation is completely lost from 50% of the normally methylated alleles of imprinted genes by E10.5, while the methylation on other genomic sequences is largely normal.

Dnmt1o is localized to the nucleus specifically at the 8-cell stage of preimplantation development. The nuclear localization of Dnmt1o only at one preimplantation stage suggests that it is only active during the 4th S-phase of preimplantation development. Loss of Dnmt1o maintenance methylation at the 4th S-phase, and subsequent maintenance methylation [by an unknown enzyme(s)] would lead to a 50% loss of methylation on imprinted genes. Notably, the loss of maintenance methylation activity in Dnmt1o-deficient embryos leads to embryos that are epigenetic mosaics, with different cells showing loss of methylation at some imprinted genes and normal methylation at others (Reinhart and Chaillet, unpublished). This finding is consistent with loss of maintenance methylation at the 4th S-phase, and suggests that methylation on the exceptional IAP element and imprinted gene DMD sequences is maintained by Dnmt1o. The experiments described in Chapter 3 and Chapter 4 analyzed the methylation of imprinted gene DMD sequences and IAP element sequences, and provide a novel approach to investigate this process.

The requirement of sequences from imprinted gene DMDs for the establishment and maintenance of parent-specific DNA methylation was examined using derivatives of the *RSVlgmyc* transgene. The *RSVlgmyc* transgene exhibits maternal-specific DMD methylation. The non-imprinted *Ig/myc* transgene was generated by removal of the *RSVlgmyc* transgene DMD. Addition of specific sequences from endogenous DMDs to the *Ig/myc* transgene restored transgene imprinting. This suggests that specific DMD sequences, even removed from their endogenous context, possess the ability to create a DMD. Importantly, this ability is only observed in combination with the *Ig/myc* sequences. By comparing the sequences that were able to restore imprinting to *Ig/myc*, to those that were not able to restore imprinting, we can draw some initial conclusions about the sequences that are required to create a DMD.

Using *Igf2r* DMD sequences (maternally methylated) we have observed that tandem repeats, specifically the TR2+3 repeats, were able to create a maternally methylated transgene locus. Similarly, hybrid transgenes containing the tandem repeats from the *Snprn* or *Kcnq1* loci (both maternally methylated) were also able to restore maternal-specific transgene methylation. This suggests a role for tandem repeats in creating an allele-specific methylation mark. Supporting this idea we have also shown that a transgene containing one unit copy of the TR2+3 repeat is not imprinted.

The ability of multiple, tandem repeat-containing sequences to restore imprinting to *Igf2r*, demonstrates that tandem repeats at endogenous loci are important to the imprinting process. How are tandem repeats involved in this process? As mentioned above, in order to be an effective imprint, DMD methylation must be specifically established in one gamete, maintained during preimplantation and throughout embryogenesis. Loss of the embryonic Dnmt1 enzyme leads to loss of all genomic methylation, including methylation at imprinted loci (Li et al. 1993). This suggests that embryonic maintenance methyltransferase activity is non-specific. In contrast, during preimplantation development, methylation is specifically maintained at imprinted loci, and loss of the Dnmt1 methyltransferase specifically affects imprints. If tandem repeats are performing an essential function, they are most likely involved in establishment of DMD methylation in the oocyte, or maintenance of DMD methylation during preimplantation. Analysis of transgene methylation during preimplantation showed that the transgene is methylated at the 4-cell, 8-cell, and blastocyst stages, supporting a role for repeats in either/or both processes.

The function of repeats at specific stages of the imprinting process may be addressed by examining allele-specific methylation on non-imprinted and imprinted transgenes in the gametes

and during preimplantation development. For example, the TR2+3 repeat is capable of imprinting the transgene in 2.5 unit copies, but not in one unit copy. If the single unit copy TR2+3 transgene is differentially methylated in the oocyte but eventually loses methylation by the adult this would demonstrate that repeats are required for maintenance methylation. However, if the single unit copy TR2+3 transgene is not differentially methylated in the gametes, then the repeat sequences are also required for methylation establishment in the oocyte. Most of the non-imprinted transgenes examined contained a relatively high level of methylation. This may suggest that tandem repeats are not only required for maintaining methylation on the methylated DMD allele, but are required for protecting the paternal allele from acquiring methylation during development. Again, this question may be addressed by examining allele-specific methylation on non-imprinted and imprinted transgenes during preimplantation and postimplantation development.

The question also arises as to whether the requirement for repeats is specific or non-specific. Will any repeat sequence, above a certain size, perform the same function? This question has been in part addressed by the fact that some repeats do not imprint the transgene (including the TR1 and IAP repeats). However, it would be a better test to generate a “random” repeated sequence, of a similar size to either the *Igf2r* or *Snrpn* repeats. The inability of a random repeat to imprint the *Ig/myc* transgene would support a role for specific repeats in genomic imprinting.

The entire 2 kb *H19* DMD (paternally methylated) was unable to restore imprinting to the *Ig/myc* transgene. The *H19* and *RSVlgmyc* loci are oppositely imprinted, and the *H19* DMD does not contain tandem repeats. This could suggest that sequences specific to each gamete are needed to create an imprint, or simply that the *H19* DMD sequences could not create an imprint

in this model system. This can be further tested by placing sequences from other paternally methylated, endogenous DMD sequences into the *Ig/myc* transgene. For example, the *Rasgrfl* locus is paternally methylated and contains a tandem repeat in its DMD that is required for imprinting (41-mer repeated 40 times) (Yoon et al. 2002). The ability or inability of these sequences to imprint the transgene could clarify these results.

Like the methylated alleles of imprinted loci, the methylation at IAPs is maintained during preimplantation development (Lane et al. 2003). However, data suggest that at the blastocyst stage a subset of IAP elements lose methylation (described in section 1.6.3). Examining the methylation of IAP LTRs at unique genomic insertion sites has provided useful information regarding the features of this methylation loss. We have shown that the methylation on the majority of IAPs is maintained at the blastocyst stage, while the methylation of specific IAPs is lost. Still, IAPs that are unmethylated in the blastocyst are targeted for *de novo* methylation later in development. The question remains as to how methylation is maintained on specific IAP element sequences during preimplantation development.

The shared maintenance of imprints and IAPs during preimplantation development indicates that a similar mechanism regulates their methylation. As mentioned above, the Dnmt1o protein may maintain methylation on both imprinted gene DMD sequences and IAP sequences during preimplantation development. IAP LTR methylation was examined by Southern blot in E10.5 Dnmt1o-deficient embryos (Howell et al. 2001). However, IAP elements may be targets for *de novo* methylation throughout embryogenesis (suggested by the remethylation of IAPs found to be unmethylated in the blastocyst), and analysis of methylation in E10.5 Dnmt1o-deficient embryos may only assess IAPs that are remethylated following implantation of the blastocyst. Specifically examining the methylation of individual, methylated

IAP insertions at the blastocyst stage of development, shortly following the activity of Dnmt1o, will hopefully determine whether common mechanisms are involved in regulating both of these processes.

We have emphasized the importance of the establishment and maintenance of methylation during early development for both imprinted gene DMDs and IAP elements. The aberrant methylation of certain sequences may also occur at these times. For example, the expanded trinucleotide repeat tract at the human *FMRI* locus is associated with a high level of DNA methylation (Pieretti et al. 1991) (described in section 1.7). The association between an expanded CGG repeat tract and methylation suggests that methylation may influence trinucleotide repeat stability in the gamete or early embryo. However, little is known about the timing of repeat expansion and methylation, or the effect of methylation on repeat stability. A mouse model has not been generated that recreates the dynamic expansions seen at the human *FMRI* locus. We attempted to create a mouse transgene with maternal-specific methylation at a CGG repeat tract (similar to the DMD of the *Ig/myc* hybrid transgenes) to test the involvement of DNA methylation in trinucleotide repeat expansion. Unfortunately, we did not observe repeat expansion using this mouse transgene. However, the transgenic lines generated were not consistently maternally methylated in adult tissues, suggesting that they may not establish or maintain methylation in the early embryo. Generating new transgenic lines, or a different mouse transgene that is methylated in the gamete and preimplantation embryo may effectively test the influence of methylation on repeat stability.

BIBLIOGRAPHY

- (1994). Fmr1 knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. Cell **78**(1): 23-33.
- Aapola, U., K. Kawasaki, H. S. Scott, J. Ollila, M. Vihinen, M. Heino, A. Shintani, S. Minoshima, K. Krohn, S. E. Antonarakis, N. Shimizu, J. Kudoh and P. Peterson (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics **65**(3): 293-8.
- Aapola, U., I. Liiv and P. Peterson (2002). Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. Nucleic Acids Res **30**(16): 3602-8.
- Aguirre-Arteta, A. M., I. Grunewald, M. C. Cardoso and H. Leonhardt (2000). Expression of an alternative Dnmt1 isoform during muscle differentiation. Cell Growth Differ **11**(10): 551-9.
- Ainscough, J. F., T. Koide, M. Tada, S. Barton and M. A. Surani (1997). Imprinting of Igf2 and H19 from a 130 kb YAC transgene. Development **124**(18): 3621-32.
- Amariglio, N. and G. Rechavi (1993). Insertional mutagenesis by transposable elements in the mammalian genome. Environ Mol Mutagen **21**(3): 212-8.
- Aota, S., T. Gojobori, K. Shigesada, H. Ozeki and T. Ikemura (1987). Nucleotide sequence and molecular evolution of mouse retrovirus-like IAP elements. Gene **56**(1): 1-12.
- Araujo, F. D., S. Croteau, A. D. Slack, S. Milutinovic, P. Bigey, G. B. Price, M. Zannis-Hajopoulos and M. Szyf (2001). The DNMT1 target recognition domain resides in the N terminus. J Biol Chem **276**(10): 6930-6.
- Ashley, C. T., Jr., K. D. Wilkinson, D. Reines and S. T. Warren (1993). FMR1 protein: conserved RNP family domains and selective RNA binding. Science **262**(5133): 563-6.
- Assaad, F. F. and E. R. Signer (1992). Somatic and germinal recombination of a direct repeat in Arabidopsis. Genetics **132**(2): 553-66.
- Balakumaran, B. S., C. H. Freudenreich and V. A. Zakian (2000). CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. Hum Mol Genet **9**(1): 93-100.

- Barahona, A. (1997). Barbara McClintock and the transposition concept. Arch Int Hist Sci (Paris) **46**(137): 309-29.
- Barbot, W., A. Dupressoir, V. Lazar and T. Heidmann (2002). Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. Nucleic Acids Res **30**(11): 2365-73.
- Bartolomei, M. S., S. Zemel and S. M. Tilghman (1991). Parental imprinting of the mouse H19 gene. Nature **351**(6322): 153-5.
- Baskaran, S., S. Datta, A. Mandal, N. Gulati, S. Totey, R. R. Anand and V. Brahmachari (2002). Instability of CGG repeats in transgenic mice. Genomics **80**(2): 151-7.
- Bell, M. V., M. C. Hirst, Y. Nakahori, R. N. MacKinnon, A. Roche, T. J. Flint, P. A. Jacobs, N. Tommerup, L. Tranebjaerg, U. Froster-Iskenius and et al. (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell **64**(4): 861-6.
- Bestor, T. H. and V. M. Ingram (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A **80**(18): 5559-63.
- Bestor, T. H. (1988). Cloning of a mammalian DNA methyltransferase. Gene **74**(1): 9-12.
- Bestor, T. H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet **9**(16): 2395-402.
- Bonfils, C., N. Beaulieu, E. Chan, J. Cotton-Montpetit and A. R. MacLeod (2000). Characterization of the human DNA methyltransferase splice variant Dnmt1b. J Biol Chem **275**(15): 10754-60.
- Bontekoe, C. J., C. E. Bakker, I. M. Nieuwenhuizen, H. van der Linde, H. Lans, D. de Lange, M. C. Hirst and B. A. Oostra (2001). Instability of a (CGG)₉₈ repeat in the Fmr1 promoter. Hum Mol Genet **10**(16): 1693-9.
- Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman and T. H. Bestor (2001). Dnmt3L and the establishment of maternal genomic imprints. Science **294**(5551): 2536-9.
- Bressler, J., T. F. Tsai, M. Y. Wu, S. F. Tsai, M. A. Ramirez, D. Armstrong and A. L. Beaudet (2001). The SNRPN promoter is not required for genomic imprinting of the Prader-Willi/Angelman domain in mice. Nat Genet **28**(3): 232-40.
- Brown, W. T., G. E. Houck, Jr., X. Ding, N. Zhong, S. Nolin, A. Glicksman, C. Dobkin and E. C. Jenkins (1996). Reverse mutations in the fragile X syndrome. Am J Med Genet **64**(2): 287-92.

- Buiting, K., S. Gross, C. Lich, G. Gillesen-Kaesbach, O. el-Maarri and B. Horsthemke (2003). Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. Am J Hum Genet **72**(3): 571-7.
- Burt, D. W., A. D. Reith and W. J. Brammar (1984). A retroviral provirus closely associated with the Ren-2 gene of DBA/2 mice. Nucleic Acids Res **12**(22): 8579-93.
- Carlone, D. L. and D. G. Skalnik (2001). CpG binding protein is crucial for early embryonic development. Mol Cell Biol **21**(22): 7601-6.
- Carlson, L. L., A. W. Page and T. H. Bestor (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev **6**(12B): 2536-41.
- Cattanach, B. M. and M. Kirk (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature **315**(6019): 496-8.
- Cattanach, B. M., J. A. Barr, E. P. Evans, M. Burtenshaw, C. V. Beechey, S. E. Leff, C. I. Brannan, N. G. Copeland, N. A. Jenkins and J. Jones (1992). A candidate mouse model for Prader-Willi syndrome which shows an absence of Snrpn expression. Nat Genet **2**(4): 270-4.
- Chaillet, J. R., T. F. Vogt, D. R. Beier and P. Leder (1991). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell **66**(1): 77-83.
- Chaillet, J. R. (1994). Genomic imprinting: lessons from mouse transgenes. Mutat Res **307**(2): 441-9.
- Chaillet, J. R., D. S. Bader and P. Leder (1995). Regulation of genomic imprinting by gametic and embryonic processes. Genes Dev **9**(10): 1177-87.
- Chang, S. and C. Staben (1994). Directed replacement of mt A by mt a-1 effects a mating type switch in *Neurospora crassa*. Genetics **138**(1): 75-81.
- Chedin, F., M. R. Lieber and C. L. Hsieh (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A **99**(26): 16916-21.
- Chen, T., Y. Ueda, S. Xie and E. Li (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem **277**(41): 38746-54.
- Cranston, M. J., T. L. Spinka, D. A. Elson and M. S. Bartolomei (2001). Elucidation of the minimal sequence required to imprint H19 transgenes. Genomics **73**(1): 98-107.

- Cross, S. H., R. R. Meehan, X. Nan and A. Bird (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet **16**(3): 256-9.
- Cummings, C. J. and H. Y. Zoghbi (2000). Fourteen and counting: unraveling trinucleotide repeat diseases. Hum Mol Genet **9**(6): 909-16.
- Deininger, P. L. and M. A. Batzer (2002). Mammalian retroelements. Genome Res **12**(10): 1455-65.
- Dennis, K., T. Fan, T. Geiman, Q. Yan and K. Muegge (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev **15**(22): 2940-4.
- Deplus, R., C. Brenner, W. A. Burgers, P. Putmans, T. Kouzarides, Y. de Launoit and F. Fuks (2002). Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucleic Acids Res **30**(17): 3831-8.
- Ding, F. and J. R. Chaillet (2002). In vivo stabilization of the Dnmt1 (cytosine-5)-methyltransferase protein. Proc Natl Acad Sci U S A **99**(23): 14861-6.
- Dupressoir, A. and T. Heidmann (1997). Expression of intracisternal A-particle retrotransposons in primary tumors of oncogene-expressing transgenic mice. Oncogene **14**(24): 2951-8.
- Eiberg, H. and J. Mohr (1996). Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. Eur J Hum Genet **4**(4): 237-41.
- Eichler, E. E. and D. L. Nelson (1996). Genetic variation and evolutionary stability of the FMR1 CGG repeat in six closed human populations. Am J Med Genet **64**(1): 220-5.
- Elson, D. A. and M. S. Bartolomei (1997). A 5' differentially methylated sequence and the 3'-flanking region are necessary for H19 transgene imprinting. Mol Cell Biol **17**(1): 309-17.
- Falzon, M. and E. L. Kuff (1991). Binding of the transcription factor EBP-80 mediates the methylation response of an intracisternal A-particle long terminal repeat promoter. Mol Cell Biol **11**(1): 117-25.
- Fatemi, M., A. Hermann, S. Pradhan and A. Jeltsch (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol **309**(5): 1189-99.
- Fatemi, M., A. Hermann, H. Gowher and A. Jeltsch (2002). Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur J Biochem **269**(20): 4981-4.
- Feenstra, A., J. Fewell, K. Lueders and E. Kuff (1986). In vitro methylation inhibits the promotor activity of a cloned intracisternal A-particle LTR. Nucleic Acids Res **14**(10): 4343-52.

- Fehrmann, F., M. Jung, R. Zimmermann and H. G. Krausslich (2003). Transport of the intracisternal A-type particle Gag polyprotein to the endoplasmic reticulum is mediated by the signal recognition particle. J Virol **77**(11): 6293-304.
- Feng, Q. and Y. Zhang (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev **15**(7): 827-32.
- Fisch, G. S., K. Snow, S. N. Thibodeau, M. Chalifaux, J. J. Holden, D. L. Nelson, P. N. Howard-Peebles and A. Maddalena (1995). The fragile X premutation in carriers and its effect on mutation size in offspring. Am J Hum Genet **56**(5): 1147-55.
- Fitzpatrick, G. V., P. D. Soloway and M. J. Higgins (2002). The brain on microarrays. Nat Genet **32**(3): 426-31.
- Fujita, N., S. Watanabe, T. Ichimura, S. Tsuruzoe, Y. Shinkai, M. Tachibana, T. Chiba and M. Nakao (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J Biol Chem **278**(26): 24132-8.
- Fuks, F., P. J. Hurd, D. Wolf, X. Nan, A. P. Bird and T. Kouzarides (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem **278**(6): 4035-40.
- Fuks, F., P. J. Hurd, R. Deplus and T. Kouzarides (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res **31**(9): 2305-12.
- Gacy, A. M., G. Goellner, N. Juranic, S. Macura and C. T. McMurray (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell **81**(4): 533-40.
- Garry, R. F., C. D. Fermin, D. J. Hart, S. S. Alexander, L. A. Donehower and H. Luo-Zhang (1990). Detection of a human intracisternal A-type retroviral particle antigenically related to HIV. Science **250**(4984): 1127-9.
- Gedeon, A. K., E. Baker, H. Robinson, M. W. Partington, B. Gross, A. Manca, B. Korn, A. Poustka, S. Yu, G. R. Sutherland and et al. (1992). Fragile X syndrome without CCG amplification has an FMR1 deletion. Nat Genet **1**(5): 341-4.
- Genc, B., H. Muller-Hartmann, M. Zeschnigk, H. Deissler, B. Schmitz, F. Majewski, A. von Gontard and W. Doerfler (2000). Methylation mosaicism of 5'-(CGG)(n)-3' repeats in fragile X, premutation and normal individuals. Nucleic Acids Res **28**(10): 2141-52.
- Gibbons, R. J., T. L. McDowell, S. Raman, D. M. O'Rourke, D. Garrick, H. Ayyub and D. R. Higgs (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet **24**(4): 368-71.

- Glenn, C. C., D. J. Driscoll, T. P. Yang and R. D. Nicholls (1997). Genomic imprinting: potential function and mechanisms revealed by the Prader-Willi and Angelman syndromes. Mol Hum Reprod **3**(4): 321-32.
- Gordenin, D. A., T. A. Kunkel and M. A. Resnick (1997). Repeat expansion--all in a flap? Nat Genet **16**(2): 116-8.
- Goto, T. and M. Monk (1998). Regulation of X-chromosome inactivation in development in mice and humans. Microbiol Mol Biol Rev **62**(2): 362-78.
- Gourdon, G., F. Radvanyi, A. S. Lia, C. Duros, M. Blanche, M. Abitbol, C. Junien and H. Hofmann-Radvanyi (1997). Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. Nat Genet **15**(2): 190-2.
- Hagerman, R. J., K. Van Housen, A. C. Smith and L. McGavran (1984). Consideration of connective tissue dysfunction in the fragile X syndrome. Am J Med Genet **17**(1): 111-21.
- Hall, I. M., G. D. Shankaranarayana, K. Noma, N. Ayoub, A. Cohen and S. I. Grewal (2002). Establishment and maintenance of a heterochromatin domain. Science **297**(5590): 2232-7.
- Hanel, M. L. and R. Wevrick (2001). Establishment and maintenance of DNA methylation patterns in mouse Ndn: implications for maintenance of imprinting in target genes of the imprinting center. Mol Cell Biol **21**(7): 2384-92.
- Hansen, R. S., C. Wijmenga, P. Luo, A. M. Stanek, T. K. Canfield, C. M. Weemaes and S. M. Gartler (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A **96**(25): 14412-7.
- Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse and S. M. Tilghman (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature **405**(6785): 486-9.
- Hata, K., M. Okano, H. Lei and E. Li (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development **129**(8): 1983-93.
- Hendrich, B., J. Guy, B. Ramsahoye, V. A. Wilson and A. Bird (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev **15**(6): 710-23.
- Hermann, A., S. Schmitt and A. Jeltsch (2003). The human Dnmt2 has residual DNA-(Cytosine-C5)-methyltransferase activity. J Biol Chem.
- Hirst, M. C. and P. J. White (1998). Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. Nucleic Acids Res **26**(10): 2353-8.

- Hornstra, I. K., D. L. Nelson, S. T. Warren and T. P. Yang (1993). High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. Hum Mol Genet **2**(10): 1659-65.
- Howell, C. Y., A. L. Steptoe, M. W. Miller and J. R. Chaillet (1998). cis-Acting signal for inheritance of imprinted DNA methylation patterns in the preimplantation mouse embryo. Mol Cell Biol **18**(7): 4149-56.
- Howell, C. Y., T. H. Bestor, F. Ding, K. E. Latham, C. Mertineit, J. M. Trasler and J. R. Chaillet (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell **104**(6): 829-38.
- Hsieh, C. L. (1999). In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. Mol Cell Biol **19**(12): 8211-8.
- Hsu, D. W., M. J. Lin, T. L. Lee, S. C. Wen, X. Chen and C. K. Shen (1999). Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. Proc Natl Acad Sci U S A **96**(17): 9751-6.
- Irelan, J. T., A. T. Hagemann and E. U. Selker (1994). High frequency repeat-induced point mutation (RIP) is not associated with efficient recombination in Neurospora. Genetics **138**(4): 1093-103.
- Iyer, R. R., A. Pluciennik, W. A. Rosche, R. R. Sinden and R. D. Wells (2000). DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in Escherichia coli. J Biol Chem **275**(3): 2174-84.
- Jin, P. and S. T. Warren (2000). Understanding the molecular basis of fragile X syndrome. Hum Mol Genet **9**(6): 901-8.
- Jones, P. L., G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis and A. P. Wolffe (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet **19**(2): 187-91.
- Jorgensen, H. F. and A. Bird (2002). MeCP2 and other methyl-CpG binding proteins. Ment Retard Dev Disabil Res Rev **8**(2): 87-93.
- Kafri, T., M. Ariel, M. Brandeis, R. Shemer, L. Urven, J. McCarrey, H. Cedar and A. Razin (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev **6**(5): 705-14.
- Kaneko-Ishino, T., Y. Kuroiwa, N. Miyoshi, T. Kohda, R. Suzuki, M. Yokoyama, S. Viville, S. C. Barton, F. Ishino and M. A. Surani (1995). Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. Nat Genet **11**(1): 52-9.

- Kantheti, P., M. E. Diaz, A. E. Peden, E. E. Seong, D. F. Dolan, M. S. Robinson, J. L. Noebels and M. L. Burmeister (2003). Genetic and phenotypic analysis of the mouse mutant mh(2J), an Ap3d allele caused by IAP element insertion. Mamm Genome **14**(3): 157-67.
- Kim, G. D., J. Ni, N. Kelesoglu, R. J. Roberts and S. Pradhan (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. Embo J **21**(15): 4183-95.
- Kim, J., A. Kollhoff, A. Bergmann and L. Stubbs (2003). Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. Hum Mol Genet **12**(3): 233-45.
- Kremer, E. J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland and R. I. Richards (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science **252**(5013): 1711-4.
- Krayer, H., M. Mila, G. Glover, P. Carbonell, F. Ballesta and X. Estivill (1994). Fragile X syndrome and the (CGG)n mutation: two families with discordant MZ twins. Am J Hum Genet **54**(3): 437-42.
- Kuff, E. L. and J. W. Fewell (1985). Intracisternal A-particle gene expression in normal mouse thymus tissue: gene products and strain-related variability. Mol Cell Biol **5**(3): 474-83.
- Kuff, E. L. and K. K. Lueders (1988). The intracisternal A-particle gene family: structure and functional aspects. Adv Cancer Res **51**: 183-276.
- Kunst, C. B. and S. T. Warren (1994). Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell **77**(6): 853-61.
- La Spada, A. R., H. L. Paulson and K. H. Fischbeck (1994). Trinucleotide repeat expansion in neurological disease. Ann Neurol **36**(6): 814-22.
- Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter and W. Reik (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis **35**(2): 88-93.
- Lavedan, C. N., L. Garrett and R. L. Nussbaum (1997). Trinucleotide repeats (CGG)22TGG(CGG)43TGG(CGG)21 from the fragile X gene remain stable in transgenic mice. Hum Genet **100**(3-4): 407-14.
- Lavedan, C., E. Grabczyk, K. Usdin and R. L. Nussbaum (1998). Long uninterrupted CGG repeats within the first exon of the human FMR1 gene are not intrinsically unstable in transgenic mice. Genomics **50**(2): 229-40.
- Ledbetter, D. H., J. T. Mascarello, V. M. Riccardi, V. D. Harper, S. D. Airhart and R. J. Strobel (1982). Chromosome 15 abnormalities and the Prader-Willi syndrome: a follow-up report of 40 cases. Am J Hum Genet **34**(2): 278-85.

- Lee, J. H., K. S. Voo and D. G. Skalnik (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. J Biol Chem **276**(48): 44669-76.
- Lee, M. P. (2003). Genome-wide analysis of epigenetics in cancer. Ann N Y Acad Sci **983**: 101-9.
- Leff, S. E., C. I. Brannan, M. L. Reed, T. Ozcelik, U. Francke, N. G. Copeland and N. A. Jenkins (1992). Maternal imprinting of the mouse Snrpn gene and conserved linkage homology with the human Prader-Willi syndrome region. Nat Genet **2**(4): 259-64.
- Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch and E. Li (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development **122**(10): 3195-205.
- Leighton, P. A., J. R. Saam, R. S. Ingram, C. L. Stewart and S. M. Tilghman (1995). An enhancer deletion affects both H19 and Igf2 expression. Genes Dev **9**(17): 2079-89.
- Leonhardt, H., A. W. Page, H. U. Weier and T. H. Bestor (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell **71**(5): 865-73.
- Li, E., T. H. Bestor and R. Jaenisch (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69**(6): 915-26.
- Li, E., C. Beard and R. Jaenisch (1993). Role for DNA methylation in genomic imprinting. Nature **366**(6453): 362-5.
- Liang, G., M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird and P. A. Jones (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol **22**(2): 480-91.
- Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Lavolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson and E. S. Lander (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet **24**(4): 381-6.
- Linn, F., I. Heidmann, H. Saedler and P. Meyer (1990). Epigenetic changes in the expression of the maize A1 gene in Petunia hybrida: role of numbers of integrated gene copies and state of methylation. Mol Gen Genet **222**(2-3): 329-36.
- Liu, K., Y. F. Wang, C. Cantemir and M. T. Muller (2003). Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and DNMT3 in mammalian cells in vivo. Mol Cell Biol **23**(8): 2709-19.
- Lubs, H. A. (1969). A marker X chromosome. Am J Hum Genet **21**(3): 231-44.
- Lucifero, D., C. Mertineit, H. J. Clarke, T. H. Bestor and J. M. Trasler (2002). Methylation dynamics of imprinted genes in mouse germ cells. Genomics **79**(4): 530-8.

- Malter, H. E., J. C. Iber, R. Willemsen, E. de Graaff, J. C. Tarleton, J. Leisti, S. T. Warren and B. A. Oostra (1997). Characterization of the full fragile X syndrome mutation in fetal gametes. Nat Genet **15**(2): 165-9.
- Mangiarini, L., K. Sathasivam, A. Mahal, R. Mott, M. Seller and G. P. Bates (1997). Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nat Genet **15**(2): 197-200.
- Mann, M. R., Y. G. Chung, L. D. Nolen, R. I. Verona, K. E. Latham and M. S. Bartolomei (2003). Disruption of Imprinted Gene Methylation and Expression in Cloned Preimplantation Stage Mouse Embryos. Biol Reprod.
- McGrath, J. and D. Solter (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37**(1): 179-83.
- Merenstein, S. A., W. E. Sobesky, A. K. Taylor, J. E. Riddle, H. X. Tran and R. J. Hagerman (1996). Molecular-clinical correlations in males with an expanded FMR1 mutation. Am J Med Genet **64**(2): 388-94.
- Mertineit, C., J. A. Yoder, T. Taketo, D. W. Laird, J. M. Trasler and T. H. Bestor (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development **125**(5): 889-97.
- Michaud, E. J., M. J. van Vugt, S. J. Bultman, H. O. Sweet, M. T. Davisson and R. P. Woychik (1994). Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev **8**(12): 1463-72.
- Mietz, J. A. and E. L. Kuff (1990). Tissue and strain-specific patterns of endogenous proviral hypomethylation analyzed by two-dimensional gel electrophoresis. Proc Natl Acad Sci U S A **87**(6): 2269-73.
- Mietz, J. A., J. W. Fewell and E. L. Kuff (1992). Selective activation of a discrete family of endogenous proviral elements in normal BALB/c lymphocytes. Mol Cell Biol **12**(1): 220-8.
- Monckton, D. G., M. I. Coolbaugh, K. T. Ashizawa, M. J. Siciliano and C. T. Caskey (1997). Hypermutable myotonic dystrophy CTG repeats in transgenic mice. Nat Genet **15**(2): 193-6.
- Monk, M., M. Boubelik and S. Lehnert (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development **99**(3): 371-82.
- Morgan, R. A. and R. C. Huang (1984). Correlation of undermethylation of intracisternal A-particle genes with expression in murine plasmacytomas but not in NIH/3T3 embryo fibroblasts. Cancer Res **44**(11): 5234-41.

- Nakao, M., S. Matsui, S. Yamamoto, K. Okumura, M. Shirakawa and N. Fujita (2001). Regulation of transcription and chromatin by methyl-CpG binding protein MBD1. Brain Dev **23 Suppl 1**: S174-6.
- Nan, X., F. J. Campoy and A. Bird (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell **88**(4): 471-81.
- Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman and A. Bird (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393**(6683): 386-9.
- Nan, X. and A. Bird (2001). The biological functions of the methyl-CpG-binding protein MeCP2 and its implication in Rett syndrome. Brain Dev **23 Suppl 1**: S32-7.
- Neumann, B., P. Kubicka and D. P. Barlow (1995). Characteristics of imprinted genes. Nat Genet **9**(1): 12-3.
- Ng, H. H., Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg and A. Bird (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet **23**(1): 58-61.
- Ng, H. H., P. Jeppesen and A. Bird (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol **20**(4): 1394-406.
- Nichol, K. and C. E. Pearson (2002). CpG methylation modifies the genetic stability of cloned repeat sequences. Genome Res **12**(8): 1246-56.
- Nicholls, R. D., J. H. Knoll, M. G. Butler, S. Karam and M. Lalande (1989). Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. Nature **342**(6247): 281-5.
- Nolin, S. L., F. A. Lewis, 3rd, L. L. Ye, G. E. Houck, Jr., A. E. Glicksman, P. Limprasert, S. Y. Li, N. Zhong, A. E. Ashley, E. Feingold, S. L. Sherman and W. T. Brown (1996). Familial transmission of the FMR1 CGG repeat. Am J Hum Genet **59**(6): 1252-61.
- Oberle, I., A. Vincent, N. Abbadi, F. Rousseau, P. E. Hupkes, M. C. Hors-Cayla, S. Gilgenkrantz, B. A. Oostra and J. L. Mandel (1991). New polymorphism and a new chromosome breakpoint establish the physical and genetic mapping of DXS369 in the DXS98-FRAXA interval. Am J Med Genet **38**(2-3): 336-42.
- Ohta, T., T. A. Gray, P. K. Rogan, K. Buiting, J. M. Gabriel, S. Saitoh, B. Muralidhar, B. Bilienska, M. Krajewska-Walasek, D. J. Driscoll, B. Horsthemke, M. G. Butler and R. D. Nicholls (1999). Imprinting-mutation mechanisms in Prader-Willi syndrome. Am J Hum Genet **64**(2): 397-413.
- Okamura, K., Y. Hagiwara-Takeuchi, T. Li, T. H. Vu, M. Hirai, M. Hattori, Y. Sakaki, A. R. Hoffman and T. Ito (2000). Comparative genome analysis of the mouse imprinted gene

- impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting. Genome Res **10**(12): 1878-89.
- Okano, M., S. Xie and E. Li (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res **26**(11): 2536-40.
- Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99**(3): 247-57.
- Ono, R., H. Shiura, H. Aburatani, T. Kohda, T. Kaneko-Ishino and F. Ishino (2003). Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. Genome Res **13**(7): 1696-705.
- Opitz, J. M., J. M. Westphal and A. Daniel (1984). Discovery of a connective tissue dysplasia in the Martin-Bell syndrome. Am J Med Genet **17**(1): 101-9.
- Oswald, J., S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik and J. Walter (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol **10**(8): 475-8.
- Pearson, C. E. and R. R. Sinden (1998). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. Curr Opin Struct Biol **8**(3): 321-30.
- Peier, A. M. and D. L. Nelson (2002). Instability of a premutation-sized CGG repeat in FMR1 YAC transgenic mice. Genomics **80**(4): 423-32.
- Perna, M. G., D. Civitareale, V. De Filippis, M. Sacco, C. Cisternino and V. Tassi (1997). Absence of mutations in the gene encoding thyroid transcription factor-1 (TTF-1) in patients with thyroid dysgenesis. Thyroid **7**(3): 377-81.
- Pieretti, M., F. P. Zhang, Y. H. Fu, S. T. Warren, B. A. Oostra, C. T. Caskey and D. L. Nelson (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. Cell **66**(4): 817-22.
- Prendergast, G. C. and E. B. Ziff (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science **251**(4990): 186-9.
- Rachmilewitz, J., R. Goshen, I. Ariel, T. Schneider, N. de Groot and A. Hochberg (1992). Parental imprinting of the human H19 gene. FEBS Lett **309**(1): 25-8.
- Radtke, F., M. Hug, O. Georgiev, K. Matsuo and W. Schaffner (1996). Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites. Biol Chem Hoppe Seyler **377**(1): 47-56.
- Reinhart, B., M. Eljanne and J. R. Chaillet (2002). Shared role for differentially methylated domains of imprinted genes. Mol Cell Biol **22**(7): 2089-98.

- Reiss, A. L., H. H. Kazazian, Jr., C. M. Krebs, A. McCaughan, C. D. Boehm, M. T. Abrams and D. L. Nelson (1994). Frequency and stability of the fragile X premutation. Hum Mol Genet **3**(3): 393-8.
- Reyniers, E., J. J. Martin, P. Cras, E. Van Marck, I. Handig, H. Z. Jorens, B. A. Oostra, R. F. Kooy and P. J. Willems (1999). Postmortem examination of two fragile X brothers with an FMR1 full mutation. Am J Med Genet **84**(3): 245-9.
- Rhee, I., K. E. Bachman, B. H. Park, K. W. Jair, R. W. Yen, K. E. Schuebel, H. Cui, A. P. Feinberg, C. Lengauer, K. W. Kinzler, S. B. Baylin and B. Vogelstein (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature **416**(6880): 552-6.
- Robertson, K. D., S. Ait-Si-Ali, T. Yokochi, P. A. Wade, P. L. Jones and A. P. Wolffe (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet **25**(3): 338-42.
- Rountree, M. R., K. E. Bachman and S. B. Baylin (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet **25**(3): 269-77.
- Sato, T., M. Oyake, K. Nakamura, K. Nakao, Y. Fukusima, O. Onodera, S. Igarashi, H. Takano, K. Kikugawa, Y. Ishida, T. Shimohata, R. Koide, T. Ikeuchi, H. Tanaka, N. Futamura, R. Matsumura, T. Takayanagi, F. Tanaka, G. Sobue, O. Komure, M. Takahashi, A. Sano, Y. Ichikawa, J. Goto, I. Kanazawa and et al. (1999). Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats comparable with those in DRPLA patients. Hum Mol Genet **8**(1): 99-106.
- Satyamoorthy, K., K. Park, M. L. Atchison and C. C. Howe (1993). The intracisternal A-particle upstream element interacts with transcription factor YY1 to activate transcription: pleiotropic effects of YY1 on distinct DNA promoter elements. Mol Cell Biol **13**(11): 6621-8.
- Savouret, C., E. Brisson, J. Essers, R. Kanaar, A. Pastink, H. te Riele, C. Junien and G. Gourdon (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. Embo J **22**(9): 2264-73.
- Schoenherr, C. J., J. M. Levorse and S. M. Tilghman (2003). CTCF maintains differential methylation at the Igf2/H19 locus. Nat Genet **33**(1): 66-9.
- Selker, E. U., D. Y. Fritz and M. J. Singer (1993). Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. Science **262**(5140): 1724-8.
- Seznec, H., A. S. Lia-Baldini, C. Duros, C. Fouquet, C. Lacroix, H. Hofmann-Radvanyi, C. Junien and G. Gourdon (2000). Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. Hum Mol Genet **9**(8): 1185-94.

- Shemer, R., Y. Birger, A. D. Riggs and A. Razin (1997). Structure of the imprinted mouse *Snrpn* gene and establishment of its parental-specific methylation pattern. Proc Natl Acad Sci U S A **94**(19): 10267-72.
- Smilnich, N. J., C. D. Day, G. V. Fitzpatrick, G. M. Caldwell, A. C. Lossie, P. R. Cooper, A. C. Smallwood, J. A. Joyce, P. N. Schofield, W. Reik, R. D. Nicholls, R. Weksberg, D. J. Driscoll, E. R. Maher, T. B. Shows and M. J. Higgins (1999). A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc Natl Acad Sci U S A **96**(14): 8064-9.
- Smith, R. J., W. Dean, G. Konfortova and G. Kelsey (2003). Identification of novel imprinted genes in a genome-wide screen for maternal methylation. Genome Res **13**(4): 558-69.
- Snow, K., L. K. Doud, R. Hagerman, R. G. Pergolizzi, S. H. Erster and S. N. Thibodeau (1993). Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. Am J Hum Genet **53**(6): 1217-28.
- Stam, M., C. Belele, J. E. Dorweiler and V. L. Chandler (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. Genes Dev **16**(15): 1906-18.
- Stoger, R., P. Kubicka, C. G. Liu, T. Kafri, A. Razin, H. Cedar and D. P. Barlow (1993). Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. Cell **73**(1): 61-71.
- Strichman-Almashanu, L. Z., R. S. Lee, P. O. Onyango, E. Perlman, F. Flam, M. B. Frieman and A. P. Feinberg (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome Res **12**(4): 543-54.
- Surani, M. A., S. C. Barton and M. L. Norris (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308**(5959): 548-50.
- Sutcliffe, J. S., D. L. Nelson, F. Zhang, M. Pieretti, C. T. Caskey, D. Saxe and S. T. Warren (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet **1**(6): 397-400.
- Swain, J. L., T. A. Stewart and P. Leder (1987). Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell **50**(5): 719-27.
- Tang, L. Y., M. N. Reddy, V. Rasheva, T. L. Lee, M. J. Lin, M. S. Hung and C. K. Shen (2003). The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. J Biol Chem.
- Tassone, F., R. J. Hagerman, L. W. Gane and A. K. Taylor (1999). Strong similarities of the FMR1 mutation in multiple tissues: postmortem studies of a male with a full mutation and a male carrier of a premutation. Am J Med Genet **84**(3): 240-4.

- Taylor, A. K., F. Tassone, P. N. Dyer, S. M. Hersch, J. B. Harris, W. T. Greenough and R. J. Hagerman (1999). Tissue heterogeneity of the FMR1 mutation in a high-functioning male with fragile X syndrome. Am J Med Genet **84**(3): 233-9.
- ten Lohuis, M., A. Muller, I. Heidmann, I. Niedenhof and P. Meyer (1995). A repetitive DNA fragment carrying a hot spot for de novo DNA methylation enhances expression variegation in tobacco and petunia. Plant J **8**(6): 919-32.
- Thorvaldsen, J. L., K. L. Duran and M. S. Bartolomei (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev **12**(23): 3693-702.
- Tremblay, K. D., J. R. Saam, R. S. Ingram, S. M. Tilghman and M. S. Bartolomei (1995). A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. Nat Genet **9**(4): 407-13.
- Tremblay, K. D., K. L. Duran and M. S. Bartolomei (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. Mol Cell Biol **17**(8): 4322-9.
- Turner, G., A. Daniel and M. Frost (1980). X-linked mental retardation, macro-orchidism, and the Xq27 fragile site. J Pediatr **96**(5): 837-41.
- Urnovitz, H. B. and W. H. Murphy (1996). Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. Clin Microbiol Rev **9**(1): 72-99.
- Van den Wyngaert, I., J. Sprengel, S. U. Kass and W. H. Luyten (1998). Cloning and analysis of a novel human putative DNA methyltransferase. FEBS Lett **426**(2): 283-9.
- Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, Y. H. Fu, D. P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria, F. P. Zhang and et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell **65**(5): 905-14.
- Vincent, A., D. Heitz, C. Petit, C. Kretz, I. Oberle and J. L. Mandel (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. Nature **349**(6310): 624-6.
- Voo, K. S., D. L. Carlone, B. M. Jacobsen, A. Flodin and D. G. Skalnik (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol **20**(6): 2108-21.
- Wade, P. A. (2001). Methyl CpG-binding proteins and transcriptional repression. Bioessays **23**(12): 1131-7.
- Walsh, C. P., J. R. Chaillet and T. H. Bestor (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet **20**(2): 116-7.

- Wang, Y., D. L. Osterbur, P. L. Megaw, G. Tosini, C. Fukuhara, C. B. Green and J. C. Besharse (2001). Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse. BMC Dev Biol **1**(1): 9.
- Warnecke, P. M., J. R. Mann, M. Frommer and S. J. Clark (1998). Bisulfite sequencing in preimplantation embryos: DNA methylation profile of the upstream region of the mouse imprinted H19 gene. Genomics **51**(2): 182-90.
- Weisenberger, D. J., M. Velicescu, M. A. Preciado-Lopez, F. A. Gonzales, Y. C. Tsai, G. Liang and P. A. Jones (2002). Identification and characterization of alternatively spliced variants of DNA methyltransferase 3a in mammalian cells. Gene **298**(1): 91-9.
- White, P. J., R. H. Borts and M. C. Hirst (1999). Stability of the human fragile X (CGG)(n) triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. Mol Cell Biol **19**(8): 5675-84.
- Wigler, M. H. (1981). The inheritance of methylation patterns in vertebrates. Cell **24**(2): 285-6.
- Wigler, M., D. Levy and M. Perucho (1981). The somatic replication of DNA methylation. Cell **24**(1): 33-40.
- Wilkinson, C. R., R. Bartlett, P. Nurse and A. P. Bird (1995). The fission yeast gene *pmt1+* encodes a DNA methyltransferase homologue. Nucleic Acids Res **23**(2): 203-10.
- Wu, M., E. M. Rinchik, E. Wilkinson and D. K. Johnson (1997). Inherited somatic mosaicism caused by an intracisternal A particle insertion in the mouse tyrosinase gene. Proc Natl Acad Sci U S A **94**(3): 890-4.
- Wujcik, K. M., R. A. Morgan and R. C. Huang (1984). Transcription of intracisternal A-particle genes in mouse myeloma and Ltk- cells. J Virol **52**(1): 29-36.
- Wutz, A., O. W. Smrzka, N. Schweifer, K. Schellander, E. F. Wagner and D. P. Barlow (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. Nature **389**(6652): 745-9.
- Wutz, A., H. C. Theussl, J. Dausman, R. Jaenisch, D. P. Barlow and E. F. Wagner (2001). Non-imprinted *Igf2r* expression decreases growth and rescues the Tme mutation in mice. Development **128**(10): 1881-7.
- Xie, S., Z. Wang, M. Okano, M. Nogami, Y. Li, W. W. He, K. Okumura and E. Li (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. Gene **236**(1): 87-95.
- Yates, P. A., R. W. Burman, P. Mummaneni, S. Krussel and M. S. Turker (1999). Tandem B1 elements located in a mouse methylation center provide a target for de novo DNA methylation. J Biol Chem **274**(51): 36357-61.

- Yatsuki, H., K. Joh, K. Higashimoto, H. Soejima, Y. Arai, Y. Wang, I. Hatada, Y. Obata, H. Morisaki, Z. Zhang, T. Nakagawachi, Y. Satoh and T. Mukai (2002). Domain regulation of imprinting cluster in Kip2/Lit1 subdomain on mouse chromosome 7F4/F5: large-scale DNA methylation analysis reveals that DMR-Lit1 is a putative imprinting control region. Genome Res **12**(12): 1860-70.
- Yoder, J. A. and T. H. Bestor (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet **7**(2): 279-84.
- Yokochi, T. and K. D. Robertson (2002). Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. J Biol Chem **277**(14): 11735-45.
- Yoon, B. J., H. Herman, A. Sikora, L. T. Smith, C. Plass and P. D. Soloway (2002). Regulation of DNA methylation of Rasgrf1. Nat Genet **30**(1): 92-6.
- Yu, S., M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, J. C. Mulley, S. T. Warren, D. Schlessinger and et al. (1991). Fragile X genotype characterized by an unstable region of DNA. Science **252**(5010): 1179-81.
- Yu, S., J. Mulley, D. Loesch, G. Turner, A. Donnelly, A. Gedeon, D. Hillen, E. Kremer, M. Lynch, M. Pritchard and et al. (1992). Fragile-X syndrome: unique genetics of the heritable unstable element. Am J Hum Genet **50**(5): 968-80.
- Zhang, Y., H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird and D. Reinberg (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev **13**(15): 1924-35.
- Zoghbi, H. Y. and H. T. Orr (1999). Polyglutamine diseases: protein cleavage and aggregation. Curr Opin Neurobiol **9**(5): 566-70.
- Zwart, R., F. Sleutels, A. Wutz, A. H. Schinkel and D. P. Barlow (2001). Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev **15**(18): 2361-6.